OM protein - protein search, using sw model

Run on: September 25, 2001, 17:34:53 ; Search time 20.66 Seconds
(without alignments)
654.785 Million cell updates/sec

Title: US-09-462-845-2
Perfect score: 3489
Sequence: 1 MKKLITADDITAIVSVTDPQ.....HPRQRIKRLNYISSWFDQHL 657

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1104.5	31.7	632	4	US-09-016-080-1	Sequence 1, Appli
2	635	18.2	622	2	US-08-664-646A-2	Sequence 2, Appli
3	635	18.2	622	2	US-09-066-285-2	Sequence 2, Appli
4	635	18.2	622	3	US-09-261-006-2	Sequence 2, Appli
5	635	18.2	622	4	US-08-951-088-2	Sequence 2, Appli
6	272	7.8	755	5	PCT-US93-07923-3	Sequence 3, Appli
7	272	7.8	759	5	PCT-US93-07923-2	Sequence 2, Appli
8	272	7.8	766	1	US-08-230-491A-3	Sequence 3, Appli
9	272	7.8	766	1	US-08-619-280A-3	Sequence 3, Appli
10	272	7.8	766	2	US-08-940-391-3	Sequence 3, Appli
11	257.5	7.4	760	1	US-08-230-491A-2	Sequence 2, Appli
12	257.5	7.4	760	1	US-08-619-280A-2	Sequence 2, Appli
13	257.5	7.4	760	2	US-08-940-391-2	Sequence 2, Appli
14	145.5	4.2	593	5	PCT-US93-07923-11	Sequence 11, Appl
15	141.5	4.1	588	1	US-07-903-466-3	Sequence 3, Appli
16	141.5	4.1	588	5	PCT-US93-05794-3	Sequence 3, Appli
17	129	3.7	305	3	US-08-965-600-1	Sequence 1, Appli
18	126.5	3.6	317	1	US-07-688-299-1	Sequence 1, Appli
19	126.5	3.6	317	1	US-07-980-517A-1	Sequence 1, Appli
20	126.5	3.6	318	1	US-07-688-299-3	Sequence 3, Appli
21	126.5	3.6	318	1	US-07-688-299-13	Sequence 13, Appl
22	115.5	3.3	302	3	US-08-965-600-3	Sequence 3, Appli
23	115	3.3	297	2	US-08-602-359A-37	Sequence 37, Appl
24	114.5	3.3	579	1	US-08-126-564A-31	Sequence 31, Appl
25	114.5	3.3	579	5	PCT-US94-09143-31	Sequence 31, Appl
26	113.5	3.3	900	2	US-08-813-940-4	Sequence 4, Appli
27	109.5	3.1	528	3	US-08-747-221B-37	Sequence 37, Appl
28	108.5	3.1	428	1	US-08-190-802A-29	Sequence 29, Appl
29	108	3.1	597	2	US-08-883-534-6	Sequence 6, Appli
30	108	3.1	597	3	US-09-204-764-6	Sequence 6, Appli
31	107	3.1	798	1	US-08-190-802A-64	Sequence 64, Appl
32	107	3.1	798	1	US-08-190-802A-68	Sequence 68, Appl
33	107	3.1	798	2	US-08-308-818-2	Sequence 2, Appli
34	104.5	3.0	626	6	5268290-2	Patent No. 5268290
35	104.5	3.0	1112	2	US-08-714-402-2	Sequence 2, Appli
36	104	3.0	606	2	US-08-883-534-3	Sequence 3, Appli
37	104	3.0	606	3	US-09-204-764-3	Sequence 3, Appli
38	104	3.0	1088	4	US-09-130-242-2	Sequence 2, Appli
39	103.5	3.0	346	2	US-08-602-359A-34	Sequence 34, Appl
40	103.5	3.0	758	1	US-08-258-188-2	Sequence 2, Appli
41	103.5	3.0	758	1	US-08-526-813-2	Sequence 2, Appli
42	103.5	3.0	758	5	PCT-US95-08554-2	Sequence 2, Appli
43	102.5	2.9	428	4	US-09-139-064-2	Sequence 2, Appli
44	102.5	2.9	428	4	US-09-139-064-4	Sequence 4, Appli
45	101.5	2.9	566	3	US-09-264-737-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kaxuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

Query Match 31.7%; Score 1104.5; DB 4; Length 632;
Best Local Similarity 39.9%; Pred. No. 1.1e-90;
Matches 255; Conservative 126; Mismatches 217; Indels 41; Gaps 21;

```
Qy      27 RAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRTLAFIS--DR 84
           :| :  :::: : | | | :::|: |     |:| | | | :||:||:|: :|| |  |:
Db      25 KAVFQVTEISLKDDDYFSKLYLYDGKR--VKPFTSGNKDS-NPRFSPNGKLIAFTSKRDK 81

Qy      85 EGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQD 144
           ||  ::||:: |:|||| |    ||:    :: ||:|| |   :       | ||   |
Db      82 EGKESELYVIPTDGGEARLLAKFKYGIKNLRFTEDGKSIAVVTPI-------DVEKKGND 134

Qy     145 SYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLV 204
                :: : :  :| |   |    : || |:||: | ||    :     |  :|: |
Db     135 DVH--IIREIPFWFNGVGWIYGKRNVVYLVDVESGKKKRLTPKNLNVDQIRFH-NGR-LY 190

Qy     205 FSANLTETDDASKP--HDVYIMSLESGDLKQVTPHRGSFGSSSFSP--DGRYLALLGNEK 260
           |:|   : |   ||   |:|:  ||:  ::::||  | :    | |  || :: |  |
Db     191 FTA---QEDRERKPLISDLYV--LENRKVRKLTP--GKWRILDFLPLDDGSFV-LKANTL 242

Qy     261 EYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVI 320
           |    | :  : || : | |  ||: || :  ::|  |  : |:::  :  |:   :||
Db     243 ERGIPTNAHIYHYDPKTGELKKLTKDLDRNAYNSLNSD--VRGSQRAELVYKEGWIYYV- 299
```

```
Qy     321 GTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIA--SVTKPDRPSELYSIPL 378
Db     300 ATDGPRANLFRVNLDGKIERVIGGDRSVESFDIG-DYIAFTAQDAVT----PTELYIYRD 354

Qy     379 GQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHG 438
Db     355 G-KEKKVTDFN-KWIKGYTLSKPEHFKVKASDGVEIDAWVMKPVNFRKGKKYPAILEIHG 412

Qy     439 GPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVD 498
Db     413 GPKTAYGYAFMHEFHVLTSKGFVVIFSNPRGSDGYGEEFAD-IRGHYGERDYQDLMEVVD 471

Qy     499 EAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYF 558
Db     472 EALRRFDFIDGERLGVTGGSYGGFMTNWIVGHTNRFKAAVTQRSISNWISFFGTTDIGYY 531

Qy     559 FTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMG 618
Db     532 FAPDQIGKDPWSNLEGYWEKSPLKYAPNVETPLLIIHSTEDYRCWLPEALQLFISLKYLG 591

Qy     619 KETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
Db     592 KRVELAIFPGENHDLSRSGKPKHRVKRLELIAGWMEKWL 630
```

```
RESULT   2
US-08-664-646A-2
; Sequence 2, Application US/08664646A
; Patent No. 5877001
;  GENERAL INFORMATION:
;    APPLICANT:  Murphy et al.
;    TITLE OF INVENTION:  Amidases
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/664,646A
;      FILING DATE:  June 17, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-53
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  622 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-08-664-646A-2
```

```
  Query Match           18.2%;  Score 635;  DB 2;  Length 622;
  Best Local Similarity  29.4%;  Pred. No. 1.6e-48;
  Matches  194;  Conservative  113;  Mismatches  260;  Indels   92;  Gaps   27;
```

```
Qy      18 DPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRT 77
Db      19 DPRIR--GNLIAYTLTKANMKDNKYESTVVVEDLETGS----RRFIENASMPRISPDGRK 72

Qy      78 LAFIS-DREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESID 136
Db      73 LAFTCFNEEKKETEIWVADIQTLSAKKVLSTK-NVRSMQWNDDSRRLLVV-----GFKRR 126

Qy     137 DREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAF 196
Db     127 DDEDFVFDDDVPVWF-------DNMGFFDGEKTTFWVLDTEAEEIIE------QFEKPRF 173

Qy     197 SPDGKW----LVFSANLTETDDASKP-----HDVYIMSLESGDLKQVTPHRGSFGSSSFS 247
Db     174 S-SGLWHGDAIVVNVPHRE---GSKPALFKFYDIVLW--KDGEEEKLF-ERVSF--EAVD 224

Qy     248 PDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTE-MLDVHLADALIGD--SLIGGA 304
Db     225 SDGKRILLRGKKK--KRFISEHDWLY-LWDGELKPIYEGPLDVWEAKLTEGKVYFLTPDA 281

Qy     305 EQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIASV 364
Db     282 GRVNLWLWDGKAERVV------TGDHWI-----------------YGLDVSDGKALLLI 317

Qy     365 TKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQM 424
Db     318 MTATRIGELYL--YDGELKQVTEYNGPIFRKLKTFEPRHFRFKSKD-LEIDGWYLRPEVK 374

Qy     425 EGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGD 484
Db     375 EEKA--PVIVFVHGGPKGMYGHRFVYEMQLMASKGYYVVFVNPRGSDGYSEDFALRVLER 432

Qy     485 YGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSIS 544
Db     433 TGLEDFEDIMNGIEEFFKLEPQADRERVGITGISYGGFMTNWALTQSDLFKAGISENGIS 492

Qy     545 NWISFHGVSDIGYFFTDWQ------LEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
Db     493 YWLTSYAFSDIGLWY-DVEVIGPNPLENENF---RKL---SPLFYAQNVKAPILLIHSLE 545

Qy     599 DDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
Db     546 DYRCPLDQSLMFYNVLKDMGKEAYIAIFKRGAHGHSVRGSPRHRPKRYRLFIEFFERKL 604
```

```
RESULT   3
US-09-066-285-2
; Sequence 2, Application US/09066285
; Patent No. 5985646
;  GENERAL INFORMATION:
;    APPLICANT:  Murphy et al.
;    TITLE OF INVENTION:  Amidases
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/066,285
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/664,646
```

```
;      FILING DATE:  June 17, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-53
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  622 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-09-066-285-2


  Query Match             18.2%;  Score 635;  DB 2;  Length 622;
  Best Local Similarity   29.4%;  Pred. No. 1.6e-48;
  Matches  194;  Conservative  113;  Mismatches  260;  Indels   92;  Gaps   27;

Qy      18 DPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRT 77
           ||:    |   ||  :: | : : | | : : : :||         : :: || |||||
Db      19 DPRIR--GNLIAYTLTKANMKDNKYESTVVVEDLETGS----RRFIENASMPRISPDGRK 72

Qy      78 LAFIS-DREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESID 136
           |||   : |    ::::   :   |:|:     |    |: |   :||      |
Db      73 LAFTCFNEEKKETEIWVADIQTLSAKKVLSTK-NVRSMQWNDDSRRLLVV-----GFKRR 126

Qy     137 DREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAF 196
           | |    |   ||         | |   |      ::  :: |: |          | |
Db     127 DDEDFVFDDDVPVWF-------DNMGFFDGEKTTFWVLDTEAEEIIE------QFEKPRF 173

Qy     197 SPDGKW----LVFSANLTETDDASKP-----HDVYIMSLESGDLKQVTPHRGSFGSSSFS 247
           |  | |    :| :    |    |||     :|: :   : |: :::   | ||   :
Db     174 S-SGLWHGDAIVVNVPHRE---GSKPALFKFYDIVLW--KDGEEEKLF-ERVSF--EAVD 224

Qy     248 PDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTE-MLDVHLADALIGD--SLIGGA 304
           ||: : | | :|  |       ||| :  | |  : |  |||  |    |    |  |
Db     225 SDGKRILLRGKKK--KRFISEHDWLY-LWDGELKPIYEGPLDVWEAKLTEGKVYFLTPDA 281

Qy     305 EQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIASV 364
           :  :|  | :   |:      || ::|                  : |   :   :  :
Db     282 GRVNLWLWDGKAERVV------TGDHWI-----------------YGLDVSDGKALLLI 317

Qy     365 TKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQM 424
             |  |||      | ||:|  |    |:     |   :: ::| : ::|| :||
Db     318 MTATRIGELYL--YDGELKQVTEYNGPIFRKLKTFEPRHFRFKSKD-LEIDGWYLRPEVK 374

Qy     425 EGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGD 484
           | :   |:|: :|||| |||| : :| |::|:||| ||::||||| || ::|   |
Db     375 EEKA--PVIVFVHGGPKGMYGHRFVYEMQLMASKGYYVVFVNPRGSDGYSEDFALRVLER 432

Qy     485 YGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSIS 544
           | :|::|:|  ::|  | :|  | :|:|:|| |||||||||| : |:: ||| :::  ||
Db     433 TGLEDFEDIMNGIEEFFKLEPQADRERVGITGISYGGFMTNWALTQSDLFKAGISENGIS 492

Qy     545 NWISFHGVSDIGYFFTDWQ------LEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
           |:: :  |||| :: | :      ||:: |    ||   ||| || ||: |:|::|
Db     493 YWLTSYAFSDIGLWY-DVEVIGPNPLENENF---RKL---SPLFYAQNVKAPILLIHSLE 545

Qy     599 DDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
           | |||::|:   :  || ||||  :  |   :|  |  | || | ||      :|:: |
Db     546 DYRCPLDQSLMFYNVLKDMGKEAYIAIFKRGAHGHSVRGSPRHRPKRYRLFIEFFERKL 604


RESULT   4
US-09-261-006-2
; Sequence 2, Application US/09261006
; Patent No. 6004796
;  GENERAL INFORMATION:
;    APPLICANT:  Murphy et al.
;    TITLE OF INVENTION:  Amidases
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/261,006
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/664,646
;      FILING DATE:  June 17, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-53
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  622 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-09-261-006-2


  Query Match             18.2%;  Score 635;  DB 3;  Length 622;
  Best Local Similarity   29.4%;  Pred. No. 1.6e-48;
  Matches  194;  Conservative  113;  Mismatches  260;  Indels   92;  Gaps   27;

Qy      18 DPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRT 77
           ||:    |   ||  :: | : : | | : : : :||         : :: || |||||
Db      19 DPRIR--GNLIAYTLTKANMKDNKYESTVVVEDLETGS----RRFIENASMPRISPDGRK 72

Qy      78 LAFIS-DREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESID 136
           |||   : |    ::::   :   |:|:     |    |: |   :||      |
Db      73 LAFTCFNEEKKETEIWVADIQTLSAKKVLSTK-NVRSMQWNDDSRRLLVV-----GFKRR 126

Qy     137 DREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAF 196
           | |    |   ||         | |   |      ::  :: |: |          | |
Db     127 DDEDFVFDDDVPVWF-------DNMGFFDGEKTTFWVLDTEAEEIIE------QFEKPRF 173

Qy     197 SPDGKW----LVFSANLTETDDASKP-----HDVYIMSLESGDLKQVTPHRGSFGSSSFS 247
           |  | |    :| :    |    |||     :|: :   : |: :::   | ||   :
Db     174 S-SGLWHGDAIVVNVPHRE---GSKPALFKFYDIVLW--KDGEEEKLF-ERVSF--EAVD 224

Qy     248 PDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTE-MLDVHLADALIGD--SLIGGA 304
           ||: : | | :|  |       ||| :  | |  : |  |||  |    |    |  |
Db     225 SDGKRILLRGKKK--KRFISEHDWLY-LWDGELKPIYEGPLDVWEAKLTEGKVYFLTPDA 281

Qy     305 EQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIASV 364
           :  :|  | :   |:      || ::|                  : |   :   :  :
Db     282 GRVNLWLWDGKAERVV------TGDHWI-----------------YGLDVSDGKALLLI 317

Qy     365 TKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQM 424
             |  |||      | ||:|  |    |:     |   :: ::| : ::|| :||
Db     318 MTATRIGELYL--YDGELKQVTEYNGPIFRKLKTFEPRHFRFKSKD-LEIDGWYLRPEVK 374

Qy     425 EGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGD 484
           | :   |:|: :|||| |||| : :| |::|:||| ||::||||| || ::|   |
```

```
Db    375 EEKA--PVIVFVHGGPKGMYGHRFVYEMQLMASKGYYVVFVNPRGSDGYSEDFALRVLER 432

Qy    485 YGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSIS 544
          | :|::|:|  ::|  | :|  | :|:|:|| |||||||| : |:: ||| :::  ||
Db    433 TGLEDFEDIMNGIEEFFKLEPQADRERVGITGISYGGFMTNWALTQSDLFKAGISENGIS 492

Qy    545 NWISFHGVSDIGYFFTDWQ------LEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
          |:: :  |||| :: | :      ||:: |    ||   ||| || ||: |:|:::|
Db    493 YWLTSYAFSDIGLWY-DVEVIGPNPLENENF---RKL---SPLFYAQNVKAPILLIHSLE 545

Qy    599 DDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
          | |||::|:   :  || ||||  :  |   :|  |  | || | ||      :|:: |
Db    546 DYRCPLDQSLMFYNVLKDMGKEAYIAIFKRGAHGHSVRGSPRHRPKRYRLFIEFFERKL 604


RESULT   5
US-08-951-088-2
; Sequence 2, Application US/08951088
; Patent No. 6136583
;  GENERAL INFORMATION:
;    APPLICANT:  Murphy et al.
;    TITLE OF INVENTION:  Amidases
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/951,088
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/664,646
;      FILING DATE:  June 17, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-53
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  622 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-08-951-088-2


  Query Match             18.2%;  Score 635;  DB 4;  Length 622;
  Best Local Similarity   29.4%;  Pred. No. 1.6e-48;
  Matches  194;  Conservative  113;  Mismatches  260;  Indels   92;  Gaps   27;

Qy     18 DPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRT 77
          ||:    |   ||  :: | : : | | : : : :||        : :: || |||||
Db     19 DPRIR--GNLIAYTLTKANMKDNKYESTVVVEDLETGS----RRFIENASMPRISPDGRK 72

Qy     78 LAFIS-DREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESID 136
          |||   : |    ::::   :  |:|:      |    |: |   :||      |
Db     73 LAFTCFNEEKKETEIWVADIQTLSAKKVLSTK-NVRSMQWNDDSRRLLVV-----GFKRR 126

Qy    137 DREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAF 196
          |  |   |  ||        |  |   |      ::  :: |: |        | |
Db    127 DDEDFVFDDDVPVWF-------DNMGFFDGEKTTFWVLDTEAEEIIE------QFEKPRF 173

Qy    197 SPDGKW----LVFSANLTETDDASKP-----HDVYIMSLESGDLKQVTPHRGSFGSSSFS 247
          |  | |    :| :    |   |||     :|: :   : |: :::   | ||   :
Db    174 S-SGLWHGDAIVVNVPHRE---GSKPALFKFYDIVLW--KDGEEEKLF-ERVSF--EAVD 224

Qy    248 PDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTE-MLDVHLADALIGD--SLIGGA 304
          ||: : | | :|  |       ||| :  | |  : |  |||  |    |   |  |
Db    225 SDGKRILLRGKKK--KRFISEHDWLY-LWDGELKPIYEGPLDVWEAKLTEGKVYFLTPDA 281

Qy    305 EQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIASV 364
          :  :|  | :   |:      || ::|                    : |   :  :
Db    282 GRVNLWLWDGKAERVV------TGDHWI-----------------YGLDVSDGKALLLI 317

Qy    365 TKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQM 424
          |  |||       | ||:|  |    |:     |   :: ::| : ::|| :||
Db    318 MTATRIGELYL--YDGELKQVTEYNGPIFRKLKTFEPRHFRFKSKD-LEIDGWYLRPEVK 374

Qy    425 EGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGD 484
          | :   |:|: :|||| |||| : :| |::|:||| ||::||||| || ::|   |
Db    375 EEKA--PVIVFVHGGPKGMYGHRFVYEMQLMASKGYYVVFVNPRGSDGYSEDFALRVLER 432

Qy    485 YGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSIS 544
          | :|::|:|  ::|  | :|  | :|:|:|| |||||||| : |:: ||| :::  ||
Db    433 TGLEDFEDIMNGIEEFFKLEPQADRERVGITGISYGGFMTNWALTQSDLFKAGISENGIS 492

Qy    545 NWISFHGVSDIGYFFTDWQ------LEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
          |:: :  |||| :: | :      ||:: |    ||   ||| || ||: |:|:::|
Db    493 YWLTSYAFSDIGLWY-DVEVIGPNPLENENF---RKL---SPLFYAQNVKAPILLIHSLE 545

Qy    599 DDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
          | |||::|:   :  || ||||  :  |   :|  |  | || | ||      :|:: |
Db    546 DYRCPLDQSLMFYNVLKDMGKEAYIAIFKRGAHGHSVRGSPRHRPKRYRLFIEFFERKL 604


RESULT   6
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
;  GENERAL INFORMATION:
;    APPLICANT:  Morimoto, Chikao
;    APPLICANT:  Schlossman, Stuart F.
;    APPLICANT:  Tanaka, Toshiaki
;    TITLE OF INVENTION:  HUMAN CD26 AND METHODS FOR USE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07923
;      FILING DATE:  19930819
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/934,162
;      FILING DATE:  21-AUG-1992
;      APPLICATION NUMBER:  07/832,211
;      FILING DATE:  06-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/055002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
```

```
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  755
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
PCT-US93-07923-3


  Query Match             7.8%;  Score 272;  DB 5;  Length 755;
  Best Local Similarity   20.4%;  Pred. No. 9e-16;
  Matches  153;  Conservative  113;  Mismatches  285;  Indels  200;  Gaps   34;

Qy      15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
           |: |   :|||       : | | : |||::  ||         | :   :  |
Db      90 SINDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNNTQWV----- 144

Qy      66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
             ||| |  ||:: : :    :|:       : ::|     || |
Db     145 ----TWSPVGHKLAYVWNND-----IYVKIEPNLPSYRITWTGKEDIIYNGITDWVYEEE 195

Qy     110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
               |   |||:|  :           :| |    |   :|
Db     196 VFSAYSALWWSPNGTFLAYA-------QFNDTEVPLIEYSFYSDESLQ------------ 236

Qy     165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
           | : | |   |:|               | :|  |: |       ||  |
Db     237 ---YPKTVRVPYPKAG--------------AVNPTVKFFV-----VNTDSLS------- 266

Qy     224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
            |: :    |:|        :|      ||       :   ||   :|  ||
Db     267 -SVTNATSIQIT------APASMLIGDHYLC------DVTWATQERISLQWLRRIQNYSV 313

Qy     273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
               ||   ||  ||     : ::        :|      | :| |   || :|  ::|
Db     314 MDICDYDESSGRWNCLVARQHIEMSTT----GWVGRFRPSEPHFTLDGNSFYKIISNEEG 369

Qy     326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
           | |  |:     :  :   | |   :|: |  ::|::  |
Db     370 YRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYT 429

Qy     367 ----------PDRPSELYSIPLGQEEK--QL------------TGANDKFVR--EHTIS 399
                   |:|  : ||:   :| |  ||              :  ||| :|  |   :
Db     430 KVTCLSCELNPER-CQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSA 488

Qy     400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
           : : :|         | :::|     |  :: |    :     |||:|::: ||      | |
Db     489 LDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADTVF 548

Qy     450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
               : :   |   : |||   | : ::|:    |  : :| ::|  :  |
Db     549 RLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMG-F 607

Qy     507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
           :| ||: : | ||||::|: ::|  :  ||   :  :| |  :  |      :|:  :
Db     608 VDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWEYYDSV------YTERYMG 661

Qy     566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
             ||    :  | :   |    :   |::||  ||    :|: |:  ||  :| : :
Db     662 LPTPEDNLDHYRNSTVMSRAENFKQVEYLLIHGTADDNVHFQQSAQISKALVDVGVDFQA 721

Qy     624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653
           : : :  | : | | |          ::|   |
Db     722 MWYTDEDHGIASSTAHQHIYTHMSHFIKQCF 752


RESULT   7
PCT-US93-07923-2
; Sequence 2, Application PC/TUS9307923
;  GENERAL INFORMATION:
;    APPLICANT:  Morimoto, Chikao
;    APPLICANT:  Schlossman, Stuart F.
;    APPLICANT:  Tanaka, Toshiaki
;    TITLE OF INVENTION:  HUMAN CD26 AND METHODS FOR USE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07923
;      FILING DATE:  19930819
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/934,162
;      FILING DATE:  21-AUG-1992
;      APPLICATION NUMBER:  07/832,211
;      FILING DATE:  06-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/055002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  759
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
PCT-US93-07923-2


  Query Match             7.8%;  Score 272;  DB 5;  Length 759;
  Best Local Similarity   20.4%;  Pred. No. 9.1e-16;
  Matches  153;  Conservative  113;  Mismatches  285;  Indels  200;  Gaps   34;

Qy      15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
           |: |   :|||       : | | : |||::  ||         | :   :  |
Db      94 SINDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNNTQWV----- 148

Qy      66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
             ||| |  ||:: : :    :|:       : ::|     || |
Db     149 ----TWSPVGHKLAYVWNND-----IYVKIEPNLPSYRITWTGKEDIIYNGITDWVYEEE 199

Qy     110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
               |   |||:|  :           :| |    |   :|
Db     200 VFSAYSALWWSPNGTFLAYA-------QFNDTEVPLIEYSFYSDESLQ------------ 240

Qy     165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
           | : | |   |:|               | :|  |: |       ||  |
Db     241 ---YPKTVRVPYPKAG--------------AVNPTVKFFV-----VNTDSLS------- 270

Qy     224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
            |: :    |:|        :|      ||       :   ||   :|  ||
Db     271 -SVTNATSIQIT------APASMLIGDHYLC------DVTWATQERISLQWLRRIQNYSV 317

Qy     273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
               ||   ||  ||     : ::        :|      | :| |   || :|  ::|
Db     318 MDICDYDESSGRWNCLVARQHIEMSTT----GWVGRFRPSEPHFTLDGNSFYKIISNEEG 373

Qy     326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
           | |  |:     :  :   | |   :|: |  ::|::  |
Db     374 YRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYT 433
```

```
Qy    367 ----------PDRPSELYSIPLGQEEK--QL------------TGANDKFVR--EHTIS 399
                  |:|  : ||:   :| |  ||            :  ||| :|  |   :
Db    434 KVTCLSCELNPER-CQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSA 492

Qy    400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
          : : :|         | :::|    |  :: |   :    |||:|::: ||       | |
Db    493 LDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADTVF 552

Qy    450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
              : : :   |   : |||   | : ::|:    |  : :| ::|  :  |
Db    553 RLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMG-F 611

Qy    507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
          :| ||: : | ||||::|: ::|  :  ||  :    :| |  :  |      :|:  :
Db    612 VDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWEYYDSV------YTERYMG 665

Qy    566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
            ||    :  | :   |    :   |::||  ||    :|: |:  ||  :| : :
Db    666 LPTPEDNLDHYRNSTVMSRAENFKQVEYLLIHGTADDNVHFQQSAQISKALVDVGVDFQA 725

Qy    624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653
          : : :  | : | | |        ::|   |
Db    726 MWYTDEDHGIASSTAHQHIYTHMSHFIKQCF 756


RESULT   8
US-08-230-491A-3
; Sequence 3, Application US/08230491A
; Patent No. 5587299
;  GENERAL INFORMATION:
;    APPLICANT:  Rettig, Wolfgang J.; Scanlan, Matthew J.;
;    APPLICANT:  Garin-Chesa, Pilar; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR
;    TITLE OF INVENTION:  FIBROBLAST ACTIVATION PROTEIN ` AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FELFE & LYNCH
;      STREET:           805 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:            NEW YORK
;      COUNTRY:          USA
;      ZIP:              10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE 3.5 inch 1.2 MB STORAGE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT - ASC II
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/230,491A
;      FILING DATE:  20-APRIL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5587299man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 330
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  766 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-230-491A-3


  Query Match             7.8%;  Score 272;  DB 1;  Length 766;
  Best Local Similarity   20.4%;  Pred. No. 9.3e-16;
  Matches  153;  Conservative  113;  Mismatches  285;  Indels  200;  Gaps   34;

Qy     15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
          |: |   :|||       : | | : |||::  ||         | :   :  |
```

```
Db    101 SINDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNNTQWV----- 155

Qy     66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
          ||| |  ||:: : :     :|:       : ::|     || |
Db    156 ----TWSPVGHKLAYVWNND-----IYVKIEPNLPSYRITWTGKEDIIYNGITDWVYEEE 206

Qy    110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
              |  |||:|  :           :| |    |  |    :|
Db    207 VFSAYSALWWSPNGTFLAYA-------QFNDTEVPLIEYSFYSDESLQ----------- 247

Qy    165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
          | : | |   |:|                | :|  |: |       ||  |
Db    248 ---YPKTVRVPYPKAG--------------AVNPTVKFFV-----VNTDSLS------- 277

Qy    224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
          |: :    |:|        :|       ||       :   ||   :|  ||
Db    278 -SVTNATSIQIT------APASMLIGDHYLC------DVTWATQERISLQWLRRIQNYSV 324

Qy    273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
              ||   ||  ||      : ::        :|      | :| |   || :|  ::|
Db    325 MDICDYDESSGRWNCLVARQHIEMSTT----GWVGRFRPSEPHFTLDGNSFYKIISNEEG 380

Qy    326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
          | |  |:     :  :   | |   :|: |  ::|::  |
Db    381 YRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYT 440

Qy    367 ----------PDRPSELYSIPLGQEEK--QL------------TGANDKFVR--EHTIS 399
                  |:|  : ||:   :| |  ||              :  ||| :|  |   :
Db    441 KVTCLSCELNPER-CQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSA 499

Qy    400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
          : : :|         | :::|    |  :: |   :    |||:|::: ||       | |
Db    500 LDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADTVF 559

Qy    450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
              : : :   |   : |||   | : ::|:    |  : :| ::|  :  |
Db    560 RLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMG-F 618

Qy    507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
          :| ||: : | ||||::|: ::|  :  ||  :    :| |  :  |      :|:  :
Db    619 VDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWEYYDSV------YTERYMG 672

Qy    566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
            ||    :  | :   |    :   |::||  ||    :|: |:  ||  :| : :
Db    673 LPTPEDNLDHYRNSTVMSRAENFKQVEYLLIHGTADDNVHFQQSAQISKALVDVGVDFQA 732

Qy    624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653
          : : :  | : | | |        ::|   |
Db    733 MWYTDEDHGIASSTAHQHIYTHMSHFIKQCF 763


RESULT   9
US-08-619-280A-3
; Sequence 3, Application US/08619280A
; Patent No. 5767242
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
;    TITLE OF INVENTION:  ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,280A
;      FILING DATE:  18-MARCH-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5767242man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  766 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-619-280A-3
```

```
  Query Match             7.8%;  Score 272;  DB 1;  Length 766;
  Best Local Similarity   20.4%;  Pred. No. 9.3e-16;
  Matches  153;  Conservative  113;  Mismatches  285;  Indels  200;  Gaps   34;

Qy      15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
           |: |   :|||       : | | : |||::  ||         | :   :  |
Db     101 SINDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNNTQWV----- 155

Qy      66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
             ||| |  ||:: : :     :|:       : ::|     || |
Db     156 ----TWSPVGHKLAYVWNND-----IYVKIEPNLPSYRITWTGKEDIIYNGITDWVYEEE 206

Qy     110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
               |   |||:|  :           :| |   |  |    :|
Db     207 VFSAYSALWWSPNGTFLAYA-------QFNDTEVPLIEYSFYSDESLQ---------- 247

Qy     165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
             | : | |   |:|               | :|  |: |       ||  |
Db     248 ---YPKTVRVPYPKAG--------------AVNPTVKFFV-----VNTDSLS------- 277

Qy     224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
           |: :    |:|        :|      ||       :   ||   :|  ||
Db     278 -SVTNATSIQIT------APASMLIGDHYLC------DVTWATQERISLQWLRRIQNYSV 324

Qy     273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
                ||  ||  ||     : ::        :|      | :| |   || :|  ::|
Db     325 MDICDYDESSGRWNCLVARQHIEMSTT----GWVGRFRPSEPHFTLDGNSFYKIISNEEG 380

Qy     326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
             | |  |:     :  :   | |   :|: |  ::|::  |
Db     381 YRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYT 440

Qy     367 ----------PDRPSELYSIPLGQEEK--QL------------TGANDKFVR--EHTIS 399
                     |:|  : ||:   :| |  ||                :  ||| :|  |   :
Db     441 KVTCLSCELNPER-CQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSA 499

Qy     400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
           : : :|         | :::|    |  :: |   :    |||:|:::  ||      | |
Db     500 LDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADTVF 559

Qy     450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
               : : :   |   : |||   | : ::|:    |  : :| ::|  :  |
Db     560 RLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMG-F 618

Qy     507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
           :| ||: : | ||||::|: ::|  :  ||  :    :| |  :  |       :|:  :
Db     619 VDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWEYYDSV------YTERYMG 672

Qy     566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
             ||    :  | :   |    :   |::||  ||    :|: |:  ||  :| : :
Db     673 LPTPEDNLDHYRNSTVMSRAENFKQVEYLLIHGTADDNVHFQQSAQISKALVDVGVDFQA 732
```

```
Qy     624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653
           : : :  | : | | |      ::|   |
Db     733 MWYTDEDHGIASSTAHQHIYTHMSHFIKQCF 763
```

```
RESULT  10
US-08-940-391-3
; Sequence 3, Application US/08940391
; Patent No. 5965373
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION
;    TITLE OF INVENTION:  PROTEIN ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/940,391
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/619,280
;      FILING DATE:  18-MARCH-1996
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5965373man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  766 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-940-391-3
```

```
  Query Match             7.8%;  Score 272;  DB 2;  Length 766;
  Best Local Similarity   20.4%;  Pred. No. 9.3e-16;
  Matches  153;  Conservative  113;  Mismatches  285;  Indels  200;  Gaps   34;

Qy      15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
           |: |   :|||       : | | : |||::  ||         | :   :  |
Db     101 SINDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNNTQWV----- 155

Qy      66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
             ||| |  ||:: : :     :|:       : ::|     || |
Db     156 ----TWSPVGHKLAYVWNND-----IYVKIEPNLPSYRITWTGKEDIIYNGITDWVYEEE 206

Qy     110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
               |   |||:|  :           :| |   |  |    :|
Db     207 VFSAYSALWWSPNGTFLAYA-------QFNDTEVPLIEYSFYSDESLQ---------- 247

Qy     165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
             | : | |   |:|               | :|  |: |       ||  |
Db     248 ---YPKTVRVPYPKAG--------------AVNPTVKFFV-----VNTDSLS------- 277

Qy     224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
```

```
Db    278 -SVTNATSIQIT------APASMLIGDHYLC------DVTWATQERISLQWLRRIQNYSV 324

Qy    273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
Db    325 MDICDYDESSGRWNCLVARQHIEMSTT----GWVGRFRPSEPHFTLDGNSFYKIISNEEG 380

Qy    326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
Db    381 YRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYT 440

Qy    367 ----;-----PDRPSELYSIPLGQEEK--QL------------TGANDKFVR--EHTIS 399
Db    441 KVTCLSCELNPER-CQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSA 499

Qy    400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
Db    500 LDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADTVF 559

Qy    450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
Db    560 RLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMG-F 618

Qy    507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
Db    619 VDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWEYYDSV------YTERYMG 672

Qy    566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
Db    673 LPTPEDNLDHYRNSTVMSRAENFKQVEYLLIHGTADDNVHFQQSAQISKALVDVGVDFQA 732

Qy    624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653
Db    733 MWYTDEDHGIASSTAHQHIYTHMSHFIKQCF 763
```

```
RESULT  11
US-08-230-491A-2
; Sequence 2, Application US/08230491A
; Patent No. 5587299
;  GENERAL INFORMATION:
;    APPLICANT:  Rettig, Wolfgang J.; Scanlan, Matthew J.;
;    APPLICANT:  Garin-Chesa, Pilar; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR
;    TITLE OF INVENTION:  FIBROBLAST ACTIVATION PROTEIN ' AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FELFE & LYNCH
;      STREET:             805 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:              NEW YORK
;      COUNTRY:            USA
;      ZIP:                10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE 3.5 inch 1.2 MB STORAGE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT - ASC II
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/230,491A
;      FILING DATE:  20-APRIL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5587299man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 330
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  760 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-230-491A-2
```

Query Match 7.4%; Score 257.5; DB 1; Length 760;
Best Local Similarity 21.2%; Pred. No. 1.8e-14;
Matches 154; Conservative 119; Mismatches 272; Indels 181; Gaps 32;

```
Qy     41 SYTSNIWIYETKTGGSVP----WTHGEKRSTDPRWSPDGRTLAFISD-----REGDAAQL 91
Db     71 SADNNIVLYNIETGQSYTILSNRTMKSVNASNYGLSPDRQFVYLESDYSKLWRYSYTATY 130

Qy     92 YIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEV 151
Db    131 YIYDLSNGEFVRGNELPRPIQYLCWSPVGSKLAYVYQ----NNIYLKQRPGDPPFQ-ITF 185

Qy    152 QGLSYKRDGK---GLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPA-----FSPDGKWL 203
Db    186 NG----RENKIFNGIPDWVYEEEML-------------------PTKYALWWSPNGKFL 221

Qy    204 VFSANLTETD----------DASKPHDVYIMSLESGDLKQVTPHRGSFGSSSFSPDGRYL 253
Db    222 AY-AEFNDKDIPVIAYSYYGDEQYPRTINIPYPKAGAKNPVV--------RIFIIDTTYP 272

Qy    254 ALLGNEKEYKNATLSK-----AWLYDIEQGRLTCLTEMLDVHLADAL---------- 295
Db    273 AYVGPQEVPVPAMIASSDYYFSWLTWVTDERV-CLQWLKRVQNVSVLSICDFREDWQTWD 331

Qy    296 -------IGDSLIGGA----EQRPIWTKDSQGFYVIGTDQ-GSTGIYYI--SIEGLVYPI 341
Db    332 CPKTQEHIEESRTGWAGGFFVSRPVFSYDAISYYKIFSDKDGYKHIHYIKDTVENAIQIT 391

Qy    342 RLEKEYINSFSLSPDE----------------------------------------QHF 360
Db    392 SGKWEAINIFRVTQDSLFYSSNEFEEYPGRRNIYRISIGSYPPSKKCVTCHLRKERCQYY 451

Qy    361 IASVTK-----------PDRP-SELY------SIPLGQEEKQLTGANDKFVREHTISIP- 401
Db    452 TASFSDYAKYYALVCYGPGIPISTLHDGRTDQEIKILEENKELENA------LKNIQLPK 505

Qy    402 EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYF--HEFQVLAAK- 458
Db    506 EEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGPCSQSVRSVFAVNWISYLASKE 565

Qy    459 GYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGS 518
Db    566 GMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVEDQITAVRKFIEMG-FIDEKRIAIWGWS 624

Qy    519 YGGFMTNW-IVGQTNRFKAAVTQRSISNWISFHGVSD---IGYFFTDWQLEHDMFEDTEK 574
Db    625 YGGYVSSLALASGTGLFKCGIAVAPVSSWEYYASVYTERFMGLPTKDDNLEH--YKNSTV 682

Qy    575 LWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLS 634
Db    683 M---ARAEYFRNVD--YLLIHGTADDNVHFQNSAQIAKALVNAQVDFQAMWYSDQNHGLS 737

Qy    635 --RTGH 638
Db    738 GLSTNH 743
```

```
RESULT  12
US-08-619-280A-2
; Sequence 2, Application US/08619280A
; Patent No. 5767242
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
;    TITLE OF INVENTION:  ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Felfe & Lynch
;       STREET:  805 Third Avenue
;       CITY:  New York City
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  PC-DOS
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/619,280A
;       FILING DATE:  18-MARCH-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/230,491
;       FILING DATE:  20-APRIL-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hanson, No. 5767242man D.
;       REGISTRATION NUMBER:  30,946
;       REFERENCE/DOCKET NUMBER:  LUD 5330.1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-3884
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  760 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
US-08-619-280A-2


  Query Match             7.4%;  Score 257.5;  DB 1;  Length 760;
  Best Local Similarity   21.2%;  Pred. No. 1.8e-14;
  Matches  154;  Conservative  119;  Mismatches  272;  Indels  181;  Gaps   32;

Qy      41 SYTSNIWIYETKTGGSVP----WTHGEKRSTDPRWSPDGRTLAFISD-----REGDAAQL 91
           |  :|| :|  :|| |       |     :::  ||| : :   ||     |    |
Db      71 SADNNIVLYNIETGQSYTILSNRTMKSVNASNYGLSPDRQFVYLESDYSKLWRYSYTATY 130

Qy      92 YIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEV 151
           ||     ||  :  ::|  :    ||| |  :          :|  :::     :: :
Db     131 YIYDLSNGEFVRGNELPRPIQYLCWSPVGSKLAYVYQ----NNIYLKQRPGDPPFQ-ITF 185

Qy     152 QGLSYKRDGK---GLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPA-----FSPDGKWL 203
           |    |: |   |:    | : :|                    |      :||:||:|
Db     186 NG----RENKIFNGIPDWVYEEEML-------------------PTKYALWWSPNGKFL 221

Qy     204 VFSANLTETD----------DASKPHDVYIMSLESGDLKQVTPHRGSFGSSSFSPDGRYL 253
           : |   : |          |   |  : |   ::|   |          |  |  |
Db     222 AY-AEFNDKDIPVIAYSYYGDEQYPRTINIPYPKAGAKNPVV--------RIFIIDTTYP 272

Qy     254 ALLGNEKEYKNATLSK-----AWLYDIEQGRLTCLTEMLDVHLADAL------------ 295
           | :| ::    | ::      :||  :   |: ||  :  |     |
Db     273 AYVGPQEVPVPAMIASSDYYFSWLTWVTDERV-CLQWLKRVQNVSVLSICDFREDWQTWD 331

Qy     296 -------IGDSLIGGA----EQRPIWTKDSQGFYVIGTDQ-GSTGIYYI--SIEGLVYPI 341
                 | :|  | |      ||::: |:  :| | :|: |   |:||  ::|  :
Db     332 CPKTQEHIEESRTGWAGGFFVSRPVFSYDAISYYKIFSDKDGYKHIHYIKDTVENAIQIT 391

Qy     342 RLEKEYINSFSLSPDE----------------------------------------QHF 360
           : | || | :: |                                          |::
Db     392 SGKWEAINIFRVTQDSLFYSSNEFEEYPGRRNIYRISIGSYPPSKKCVTCHLRKERCQYY 451

Qy     361 IASVTK-----------PDRP-SELY------SIPLGQEEKQLTGANDKFVREHTISIP- 401
           || :             |  | | |:       | : :| |:|  |         | :|
Db     452 TASFSDYAKYYALVCYGPGIPISTLHDGRTDQEIKILEENKELENA------LKNIQLPK 505

Qy     402 EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYF--HEFQVLAAK- 458
           |||:    | : :   :: | | :    |||:: ::|||      : |  :    ||:|
Db     506 EEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGPCSQSVRSVFAVNWISYLASKE 565

Qy     459 GYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGS 518
           |  :  :: ||:   | : : ||    |  : :| : || : |:    || ||: : | |
Db     566 GMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVEDQITAVRKFIEMG-FIDEKRIAIWGWS 624

Qy     519 YGGFMTNW-IVGQTNRFKAAVTQRSISNWISFHGVSD---IGYFFTDWQLEHDMFEDTEK 574
           |||:::  :   |  ||  :   :|:|  :  |     :|    |  |||  ::::
Db     625 YGGYVSSLALASGTGLFKCGIAVAPVSSWEYYASVYTERFMGLPTKDDNLEH--YKNSTV 682

Qy     575 LWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLS 634
           :   :  :|  ||:   |::||  ||    : : |:  ||      : : : : :| ||
Db     683 M---ARAEYFRNVD--YLLIHGTADDNVHFQNSAQIAKALVNAQVDFQAMWYSDQNHGLS 737

Qy     635 --RTGH 638
               | |
Db     738 GLSTNH 743


RESULT  13
US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION
;    TITLE OF INVENTION:  PROTEIN ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/940,391
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/619,280
;      FILING DATE:  18-MARCH-1996
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5965373man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  760 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-940-391-2


  Query Match             7.4%;  Score 257.5;  DB 2;  Length 760;
  Best Local Similarity   21.2%;  Pred. No. 1.8e-14;
  Matches  154;  Conservative  119;  Mismatches  272;  Indels  181;  Gaps   32;

Qy      41 SYTSNIWIYETKTGGSVP----WTHGEKRSTDPRWSPDGRTLAFISD-----REGDAAQL 91
           |  :|| :|  :|| |       |     :::  ||| : :   ||     |    |
Db      71 SADNNIVLYNIETGQSYTILSNRTMKSVNASNYGLSPDRQFVYLESDYSKLWRYSYTATY 130

Qy      92 YIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEV 151
           ||     ||  :  ::|  :    ||| |  :          :|  :::     :: :
```

```
Db    131 YIYDLSNGEFVRGNELPRPIQYLCWSPVGSKLAYVYQ----NNIYLKQRPGDPPFQ-ITF 185

Qy    152 QGLSYKRDGK---GLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPA-----FSPDGKWL 203
Db    186 NG----RENKIFNGIPDWVYEEEML-------------------PTKYALWWSPNGKFL 221

Qy    204 VFSANLTETD----------DASKPHDVYIMSLESGDLKQVTPHRGSFGSSSFSPDGRYL 253
Db    222 AY-AEFNDKDIPVIAYSYYGDEQYPRTINIPYPKAGAKNPVV--------RIFIIDTTYP 272

Qy    254 ALLGNEKEYKNATLSK-----AWLYDIEQGRLTCLTEMLDVHLADAL------------- 295
Db    273 AYVGPQEVPVPAMIASSDYYFSWLTWVTDERV-CLQWLKRVQNVSVLSICDFREDWQTWD 331

Qy    296 -------IGDSLIGGA----EQRPIWTKDSQGFYVIGTDQ-GSTGIYYI--SIEGLVYPI 341
Db    332 CPKTQEHIEESRTGWAGGFFVSRPVFSYDAISYYKIFSDKDGYKHIHYIKDTVENAIQIT 391

Qy    342 RLEKEYINSFSLSPDE-------------------------------------------QHF 360
Db    392 SGKWEAINIFRVTQDSLFYSSNEFEEYPGRRNIYRISIGSYPPSKKCVTCHLRKERCQYY 451

Qy    361 IASVTK-----------PDRP-SELY------SIPLGQEEKQLTGANDKFVREHTISIP- 401
Db    452 TASFSDYAKYYALVCYGPGIPISTLHDGRTDQEIKILEENKELENA------LKNIQLPK 505

Qy    402 EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYF--HEFQVLAAK- 458
Db    506 EEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGPCSQSVRSVFAVNWISYLASKE 565

Qy    459 GYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGS 518
Db    566 GMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVEDQITAVRKFIEMG-FIDEKRIAIWGWS 624

Qy    519 YGGFMTNW-IVGQTNRFKAAVTQRSISNWISFHGVSD---IGYFFTDWQLEHDMFEDTEK 574
Db    625 YGGYVSSLALASGTGLFKCGIAVAPVSSWEYYASVYTERFMGLPTKDDNLEH--YKNSTV 682

Qy    575 LWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLS 634
Db    683 M---ARAEYFRNVD--YLLIHGTADDNVHFQNSAQIAKALVNAQVDFQAMWYSDQNHGLS 737

Qy    635 --RTGH 638
Db    738 GLSTNH 743


RESULT  14
PCT-US93-07923-11
; Sequence 11, Application PC/TUS9307923
;  GENERAL INFORMATION:
;     APPLICANT:  Morimoto, Chikao
;     APPLICANT:  Schlossman, Stuart F.
;     APPLICANT:  Tanaka, Toshiaki
;     TITLE OF INVENTION:  HUMAN CD26 AND METHODS FOR USE
;     NUMBER OF SEQUENCES:  16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;       SOFTWARE:  WordPerfect (Version 5.0)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/07923
;       FILING DATE:  19930819
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/934,162
;       FILING DATE:  21-AUG-1992
;       APPLICATION NUMBER:  07/832,211
;       FILING DATE:  06-FEB-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fraser, Janis K.
;       REGISTRATION NUMBER:  34,819
;       REFERENCE/DOCKET NUMBER:  00530/055002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  593
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
PCT-US93-07923-11


  Query Match             4.2%;  Score 145.5;  DB 5;  Length 593;
  Best Local Similarity   19.7%;  Pred. No. 0.00014;
  Matches  113;  Conservative   75;  Mismatches  197;  Indels  189;  Gaps   29;

Qy     15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
Db    101 SINDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNNTQWV----- 155

Qy     66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
Db    156 ----TWSPVGHKLAYVWNND-----IYVKIEPNLPSYRITWTGKEDIIYNGITDWVYEEE 206

Qy    110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
Db    207 VFSAYSALWWSPNGTFLAYA-------QFNDTEVPLIEYSFYSDESLQ------------ 247

Qy    165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
Db    248 ---YPKTVRVPYPKAG--------------AVNPTVKFFV-----VNTDSLS------- 277

Qy    224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
Db    278 -SVTNATSIQIT------APASMLIGDHYLC------DVTWATQERISLQWLRRIQNYSV 324

Qy    273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
Db    325 MDICDYDESSGRWNCLVARQHIEMSTT----GWVGRFRPSEPHFTLDGNSFYKIISNEEG 380

Qy    326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
Db    381 YRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYT 440

Qy    367 ----------PDRPSELYSIPLGQEEK--QL------------TGANDKFVR--EHTIS 399
Db    441 KVTCLSCELNPER-CQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSA 499

Qy    400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
Db    500 LDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADTVF 559

Qy    450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNA 480
Db    560 RLNWATYLASTENIIVASFDGRGSGYQGDKIMHA 593


RESULT  15
US-07-903-466-3
; Sequence 3, Application US/07903466
; Patent No. 5395767
;  GENERAL INFORMATION:
;     APPLICANT:  Murnane, John P.
;     APPLICANT:  Painter, Robert B.
```

; APPLICANT: Kapp, Leon N.
; APPLICANT: Yu, Loh C.
; TITLE OF INVENTION: Gene for Ataxia-Telangiectasia
; TITLE OF INVENTION: Complmentation Group D (ATDC)
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: Steuart Street Tower, 18th Fl., One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: San Francisco
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/903,466
; FILING DATE: 19920622
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: 91-077-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-777-9275
; TELEFAX: 415-543-4219
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 588 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-07-903-466-3

Query Match 4.1%; Score 141.5; DB 1; Length 588;
Best Local Similarity 19.3%; Pred. No. 0.00032;
Matches 138; Conservative 75; Mismatches 212; Indels 291; Gaps 32;

```
Qy     51 TKTGGSVPWTHGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYG 110
Db      7 SRSNGSSP------EARDAR-SPSGPSGSLENGTKADGKDAKTTNGHGGEA--------- 50

Qy    111 VSKPLWSPDGESILVTISLGEG-----------------ESIDDREKT--EQDSYEPVE 150
Db     51 -------AEGKSLGSALKPGEGRSALFAGNEWRRPIIQFVESGDDKNSNYFSMDSME--- 100

Qy    151 VQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLT 210
Db    101 ---------GK---RSPYAGLQL-----GAAKKPPVTFAEKGDVRKS------IFS---- 133

Qy    211 ETDDASKPHDVYIMSLESGDLKQVTPHRGSFGSSSFSPDGRYLAL----LGN-EKEYKNA 265
Db    134 ---ESRKP---TVSIMEPGETRRNSYPRADTGLFSRSKSGSEEVLCDSCIGNKQKAVKSC 187

Qy    266 TLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSL--------IGGAEQR--PIWTKDSQ 315
Db    188 --------------LVCQASFCELHLKPHLEGAAFRDHQLLEPIRDFEARKCPVHGKTME 233

Qy    316 GFYVIGTDQGSTGIYY-----------------------------------ISIEGLV 338
Db    234 LF--CQTDQ--TCICYLCMFQEHKNHSTVTVEEAKAEKETELSLQKEQLQLKIIEIEDEA 289

Qy    339 YPIRLEKEYINSFSLSPD---EQHFIASVTKPDRPSELYSIPLGQEEKQLTGA------- 388
Db    290 EKWQKEKDRIKSFTTNEKAILEQNFRDLVRDLEKQKEEVRAALEQREQDAVDQVKVIMDA 349

Qy    389 ---------NDKFVREHTISIPEEIQYATEDGVMVNGWLMRP------AQMEGE------ 427
Db    350 LDERAKVLHEDKQTREQLHSISDSVLFLQEFGALMSNYSLPPPLPTYHVLLEGEGLGQSL 409

Qy    428 ------------------TTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRG 469
Db    410 GNFKDDLLNVCMRHVEKMCKADLSRNFIERNHMENG----------------------G 446

Qy    470 SHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKR-DPHIDPKRLGVTGGSYGGFMTNWIV 528
Db    447 DHRYVNNYTNSFGGEWSAPD----------TMKRYSMYLTPK---------GGVRTSYQP 487

Qy    529 GQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVE 588
Db    488 SSPGRFTKETTQKNFNNLYGTKG-----------------NYTSRVWE-----YSSSIQ 524

Qy    589 TPLLILHGERDDRCPIEQAEQLFIALK-----------------KMGKETKLVRF 626
Db    525 --------NSDNDLPVVQGSSSF-SLKGYPSLMRSQSPKAQPQTWKSGKQTMLSHY 571
```

Search completed: September 25, 2001, 17:36:26
Job time: 93 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2001, 17:35:58 ; Search time 42.42 Seconds
(without alignments)
2049.137 Million cell updates/sec

Title: US-09-462-845-2
Perfect score: 3489
Sequence: 1 MKKLITADDITAIVSVTDPQ.....HPRQRIKRLNYISSWFDQHL 657

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1104.5	31.7	642	1	O58593	O58593 pyrococcus
2	1096.5	31.4	631	1	Q9uyb0	Q9uyb0 pyrococcus
3	995	28.5	655	2	Q9RXY9	Q9rxy9 deinococcus
4	962.5	27.6	659	1	Q9YAW3	Q9yaw3 aeropyrum p
5	903	25.9	674	1	Q9HN07	Q9hn07 halobacteri
6	697	20.0	572	1	Q9Y9J8	Q9y9j8 aeropyrum p
7	667.5	19.1	632	1	Q9UYQ8	Q9uyq8 pyrococcus
8	660	18.9	622	1	O58323	O58323 pyrococcus
9	473	13.6	709	2	Q9PB84	Q9pb84 xylella fas
10	439	12.6	663	5	Q9U592	Q9u592 entamoeba d
11	433.5	12.4	665	5	Q9U593	Q9u593 entamoeba h
12	410	11.8	591	1	Q9Y944	Q9y944 aeropyrum p
13	405.5	11.6	721	3	O13479	O13479 aspergillus
14	402.5	11.5	726	3	Q9UW98	Q9uw98 trichophyto
15	392.5	11.2	725	3	Q9Y8E3	Q9y8e3 aspergillus
16	391	11.2	732	6	Q9TS46	Q9ts46 sus scrofa
17	387.5	11.1	583	1	Q9HKD7	Q9hkd7 thermoplasm
18	383	11.0	732	4	Q9P0Y2	Q9p0y2 homo sapien
19	361	10.3	683	3	Q9P778	Q9p778 schizosacch
20	357.5	10.2	741	2	P95782	P95782 xanthomonas
21	355.5	10.2	598	1	Q9HS46	Q9hs46 halobacteri
22	350.5	10.0	745	5	Q9TYX1	Q9tyx1 caenorhabdi
23	346	9.9	637	2	Q55413	Q55413 synechocyst
24	334.5	9.6	606	2	Q9XAJ2	Q9xaj2 streptomyce
25	328.5	9.4	795	2	Q9PHC9	Q9phc9 xylella fas
26	319.5	9.2	569	1	P95915	P95915 sulfolobus
27	303	8.7	629	5	Q19086	Q19086 caenorhabdi
28	287.5	8.2	582	1	Q9YBQ2	Q9ybq2 aeropyrum p
29	278	8.0	711	2	Q47900	Q47900 flavobacter
30	278	8.0	765	6	Q9N2I7	Q9n2i7 felis silve
31	267.5	7.7	761	11	P97321	P97321 mus musculu
32	263	7.5	707	2	Q9F348	Q9f348 streptomyce
33	263	7.5	1042	5	Q9VC20	Q9vc20 drosophila
34	263	7.5	1102	5	Q9VC19	Q9vc19 drosophila
35	259.5	7.4	723	2	O31048	O31048 porphyromon
36	258.5	7.4	678	10	Q9FG66	Q9fg66 arabidopsis
37	257.5	7.4	760	4	O00199	O00199 homo sapien
38	257.5	7.4	760	4	Q12884	Q12884 homo sapien
39	255.5	7.3	237	2	O87543	O87543 capnocytoph
40	253.5	7.3	723	2	O66223	O66223 porphyromon
41	249.5	7.2	882	4	Q9HBM5	Q9hbm5 homo sapien
42	249	7.1	748	13	P70092	P70092 xenopus lae
43	247.5	7.1	755	13	Q91651	Q91651 xenopus lae
44	241.5	6.9	709	11	Q9QWW2	Q9qww2 mus musculu
45	240	6.9	829	5	Q18253	Q18253 caenorhabdi

ALIGNMENTS

RESULT 1
O58593
ID O58593 PRELIMINARY; PRT; 642 AA.
AC O58593;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 73.5 KDA PROTEIN PH0863.
GN PH0863.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000003; BAA29957.1; -.
DR InterPro; IPR000379; -.
KW Hypothetical protein.
SQ SEQUENCE 642 AA; 73491 MW; E9D3D28E92D754BD CRC64;

Query Match 31.7%; Score 1104.5; DB 1; Length 642;
Best Local Similarity 39.9%; Pred. No. 3.3e-68;
Matches 255; Conservative 126; Mismatches 217; Indels 41; Gaps 21;

```
Qy      27 RAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRTLAFIS--DR 84
           :| :  :::: : | | | :::|: |     |:| | | | :||:||:|: :|| |  |:
Db      35 KAVFQVTEISLKDDDYFSKLYLYDGKR--VKPFTSGNKDS-NPRFSPNGKLIAFTSKRDK 91

Qy      85 EGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQD 144
           ||  ::||:: |:|||||| |    ||:    :: ||:|| |   :       | ||   |
Db      92 EGKESELYVIPTDGGEARLLAKFKYGIKNLRFTEDGKSIAVVTPI-------DVEKKGND 144
```

20	357.5	10.2	741	2	P95782	P95782 xanthomonas
21	355.5	10.2	598	1	Q9HS46	Q9hs46 halobacteri
22	350.5	10.0	745	5	Q9TYX1	Q9tyx1 caenorhabdi
23	346	9.9	637	2	Q55413	Q55413 synechocyst
24	334.5	9.6	606	2	Q9XAJ2	Q9xaj2 streptomyce
25	328.5	9.4	795	2	Q9PHC9	Q9phc9 xylella fas
26	319.5	9.2	569	1	P95915	P95915 sulfolobus
27	303	8.7	629	5	Q19086	Q19086 caenorhabdi
28	287.5	8.2	582	1	Q9YBQ2	Q9ybq2 aeropyrum p
29	278	8.0	711	2	Q47900	Q47900 flavobacter
30	278	8.0	765	6	Q9N2I7	Q9n2i7 felis silve
31	267.5	7.7	761	11	P97321	P97321 mus musculu
32	263	7.5	707	2	Q9F348	Q9f348 streptomyce
33	263	7.5	1042	5	Q9VC20	Q9vc20 drosophila
34	263	7.5	1102	5	Q9VC19	Q9vc19 drosophila
35	259.5	7.4	723	2	O31048	O31048 porphyromon
36	258.5	7.4	678	10	Q9FG66	Q9fg66 arabidopsis
37	257.5	7.4	760	4	O00199	O00199 homo sapien
38	257.5	7.4	760	4	Q12884	Q12884 homo sapien
39	255.5	7.3	237	2	O87543	O87543 capnocytoph
40	253.5	7.3	723	2	O66223	O66223 porphyromon
41	249.5	7.2	882	4	Q9HBM5	Q9hbm5 homo sapien
42	249	7.1	748	13	P70092	P70092 xenopus lae
43	247.5	7.1	755	13	Q91651	Q91651 xenopus lae
44	241.5	6.9	709	11	Q9QWW2	Q9qww2 mus musculu
45	240	6.9	829	5	Q18253	Q18253 caenorhabdi

```
Qy    145 SYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLV 204
          :: : :  :| |   |    : || |:||: | ||    :     |  :|: |
Db    145 DVH--IIREIPFWFNGVGWIYGKRNVVYLVDVESGKKKRLTPKNLNVDQIRFH-NGR-LY 200

Qy    205 FSANLTETDDASKP--HDVYIMSLESGDLKQVTPHRGSFGSSSFSP--DGRYLALLGNEK 260
          |:|   : |   ||   |:|:  ||:  ::::||  | :    | |  || :: |  |
Db    201 FTA---QEDRERKPLISDLYV--LENRKVRKLTP--GKWRILDFLPLDDGSFV-LKANTL 252

Qy    261 EYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVI 320
          |    | :  : || : | |  ||: || :  ::|  |  : |:::  :  |:   :||
Db    253 ERGIPTNAHIYHYDPKTGELKKLTKDLDRNAYNSLNSD--VRGSQRAELVYKEGWIYYV- 309

Qy    321 GTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIA--SVTKPDRPSELYSIPL 378
           ||     :: ::::| :  :      : || :  |   | |  :||    |:|||
Db    310 ATDGPRANLFRVNLDGKIERVIGGDRSVESFDIG-DYIAFTAQDAVT----PTELYIYRD 364

Qy    379 GQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHG 438
          | :||::|  | |::: :|:| ||  :    ||| :: |:|:|          || || |||
Db    365 G-KEKKVTDFN-KWIKGYTLSKPEHFKVKASDGVEIDAWVMKPVNFRKGKKYPAILEIHG 422

Qy    439 GPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVD 498
          ||   ||: : ||| || :||: |:: ||||| |||:|| : :|| || :|| |:|: ||
Db    423 GPKTAYGYAFMHEFHVLTSKGFVVIFSNPRGSDGYGEEFAD-IRGHYGERDYQDLMEVVD 481

Qy    499 EAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYF 558
          ||::|   || :|||||||||||||||||| |||||||||||||||||| | :||||:
Db    482 EALRRFDFIDGERLGVTGGSYGGFMTNWIVGHTNRFKAAVTQRSISNWISFFGTTDIGYY 541

Qy    559 FTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMG 618
          |   |:  | : : |  |::|||||| ||||||||:|   | || : :| ||||:|| :|
Db    542 FAPDQIGKDPWSNLEGYWEKSPLKYAPNVETPLLIIHSTEDYRCWLPEALQLFISLKYLG 601

Qy    619 KETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
          |  :|  ||  :|:|||:| |: |:|||  |: | :: |
Db    602 KRVELAIFPGENHDLSRSGKPKHRVKRLELIAGWMEKWL 640


RESULT   2
Q9UYB0
ID   Q9UYB0      PRELIMINARY;      PRT;   631 AA.
AC   Q9UYB0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ACYLAMINO-ACID-RELEASING ENZYME (EC 3.4.19.1) (ACYL-PEPTIDE HYDROLASE)
DE   (APH) (ACYLAMINOACYL-PEPTIDASE).
GN   PAB1300.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORSAY;
RA   Heilig R.;
RT   "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT   structure and evolution.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ248288; CAB50502.1; -.
DR   InterPro; IPR000379; -.
SQ   SEQUENCE   631 AA;  72567 MW;  46539D5584F75E49 CRC64;


  Query Match           31.4%;  Score 1096.5;  DB 1;  Length 631;
  Best Local Similarity   38.2%;  Pred. No. 1.1e-67;
  Matches  258;  Conservative  119;  Mismatches  233;  Indels   65;  Gaps   20;

Qy      1 MKKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSV-PW 59
          | | |:  ||     | :      | :  :  :::: ::: | |:|::|:   |  |  :
Db      1 MMKKISEKDIEKFKLVGN--LDASGKKVVFQVTEISIKENDYFSSIYLYD---GRKVRRF 55

Qy     60 THGEKRSTDPRWSPDGRTLAFISDR--EGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWS 117
          | | |   :||:||||| :|| | |  ||  ::||:|  :||||| |    ||:    :|
Db     56 TFGRK-DANPRFSPDGRLIAFTSKRSKEGRESELYLMPVDGGEARLLAKFKYGIKDLRFS 114

Qy    118 PDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRD------------GKGLT 164
          ||: | |                        ||||: :: :::             | |
Db    115 EDGKEIAV--------------------ITPVEVKKVANEKEDVHVIDEIPFWFNGIGWI 154

Qy    165 RGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKP--HDVY 222
          |    | || |:||: ::::    |     |  :||  :    | : |   ||   |||
Db    155 YGRRNALYLVDVESGKKRKVSPRNLDVLSVRFH-NGKLYL----LAQEDREKKPMISDVY 209

Qy    223 IMSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTC 282
          :  ||   :|::||  | :    | |    | |  |  |   || :  :||    | |
Db    210 V--LEGKRVKKLTP--GKWRILDFVPLSDGLILKANTLERGFATTTSIYLY--RNGELRR 263

Qy    283 LTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIR 342
          | :  |:   ::|  |  : ||::  :  ||    | : ||     :: ::::| |  :
Db    264 LFKE-DLSAYNSLNSD--VRGAQRAELIFKDG-WIYFVATDGPRANLFRVNLDGKVERVI 319

Qy    343 LEKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPE 402
             : |||:     :   :      |:|||::  | :||::|  |  ::|:: :| ||
Db    320 RGDRSVESFSVG---DYIAFTAQDATTPTELYALRDG-KEKRITDFN-AWIRDYKLSKPE 374

Qy    403 EIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAV 462
             :    ||  :: |:|:|   :    || :| |||||   ||: : |||||| |||: |
Db    375 HFRVTASDGKEIDAWIMKPVDFKPGRKYPAVLEIHGGPKTAYGYAFMHEFQVLVAKGFVV 434

Qy    463 VYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGF 522
          :: ||||| |||:|| : :|| || :|| |:|: ||||:|   || :||||||||||||
Db    435 IFSNPRGSDGYGEEFAD-IRGHYGERDYQDLMEVVDEALKRFDFIDEERLGVTGGSYGGF 493

Qy    523 MTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLK 582
          |||||| ||||||||||||||||||| | :||||:|   |:  | : : |  |::||||
Db    494 MTNWIVGHTNRFKAAVTQRSISNWISFFGTTDIGYYFAPDQIGGDPWSNLEGYWEKSPLK 553

Qy    583 YAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQR 642
          || |||||||:|   | || : :| | :|||| :||  :|  ||  :|:|||:| |: |
Db    554 YAPNVETPLLIIHSTEDYRCWLPEALQFYIALKYLGKTVELAIFPGENHDLSRSGRPKHR 613

Qy    643 IKRLNYISSWFDQHL 657
          :|||  |  ||:: |
Db    614 VKRLELIVGWFERWL 628


RESULT   3
Q9RXY9
ID   Q9RXY9      PRELIMINARY;      PRT;   655 AA.
AC   Q9RXY9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ACYL-PEPTIDE HYDROLASE, PUTATIVE.
GN   DR0165.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001879; AAF09754.1; -.
DR   TIGR; DR0165; -.
DR   InterPro; IPR000379; -.
KW   Hydrolase.
```

```
SQ   SEQUENCE   655 AA;  72626 MW;  42CCCBA0FB72F200 CRC64;


  Query Match             28.5%;  Score 995;  DB 2;  Length 655;
  Best Local Similarity   36.0%;  Pred. No. 1.3e-60;
  Matches  250;  Conservative  102;  Mismatches  248;  Indels   94;  Gaps   19;

Qy       8 DDITAIVSVTDPQYAPDGTRAAYVKSQVNQE----------KDSYTSNIWIYETKTGGSV 57
           | : |:   :||| :||| : |:| :|:::|          :  | | :|: |   | :
Db      12 DSLLALAFPSDPQVSPDGKQVAFVLAQISEEDPAKPDKDFARPRYRSGLWLSE--GGAAR 69

Qy      58 PWTHGEKRSTD--PRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPL 115
           | || |    |  |||||||: |||:       | | ::  :|||||::|    ||| |
Db      70 PLTHAETGRGDSAPRWSPDGQNLAFVRSAGEVKAALMLLPLKGGEARRVTHFKNGVSGPQ 129

Qy     116 WSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKG----LTRGAY--- 168
           |||||  |  |               |  |: :         ||| :|    |||  |
Db     130 WSPDGRFIAFT--------------TTADTED---------KRDERGEARVLTRPVYRAN 166

Qy     169 ---------AQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPH 219
                  | | |  |:: :::|  : :   |  :: || : ::   :  | :
Db     167 GADWLPERPAALWLYDVEADKLREWYAPEIGIGALSWWPDSRGVLIVQSEDEWQASQWRQ 226

Qy     220 DVYIMSLESGDLKQVTPHRGSFGSSSFS----PDGRYLALLGNEKEYKNATLSKAWLYDI 275
           |||  : | : |          : |::      |||:  ||:|       |     | || |
Db     227 DVYDLPLPTADAPAAPQKLLDWNSAAHGLAPHPDGQRFALIGRPAGKGNT--EHAHLYLI 284

Qy     276 EQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIE 335
           | |:   |    |  : ||: ||  :|   : | |  |         | :|| |::   |
Db     285 ENGQHRRLDTGHDHPVGDAVGGDCHVGAFPEGPRWL-DGDTLLFSSTVRGSVGLFTAHIG 343

Qy     336 GLVYPIRLEKEY-------INSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGA 388
           | |       | |       |::|: :   :| :| :    | |       :    :::|
Db     344 GGV------KAYDHDPQGVISAFTAN---EHGVALI----RESATRFPEVELNGQRVTDL 390

Qy     389 NDKF---VREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYG 445
           : :|   |||     |: : : || |    ||:: |   |||   | :|||||||  ||
Db     391 HARFPFPVRE-----PQRVTFETELG-EGEGWVLLP---EGEQKVPALLNIHGGPHTDYG 441

Qy     446 HTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDP 505
           | : ||||::||:|| | | ||||| |||| :|:|: | :|  | ||::   |  ::  |
Db     442 HGFTHEFQLMAARGYGVCYSNPRGSVGYGQAWVDAIYGRWGTVDADDLLNFFDRCLEAVP 501

Qy     506 HIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
            :|  :  | ||||||||||||| | | ||:||:| | ||| ||| | ||||  | | :|
Db     502 RLDAAKTAVMGGSYGGFMTNWITGHTTRFQAAITDRCISNLISFGGTSDIGLRFWDDELG 561

Qy     566 HDMFE--DTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
             |     |  |||| |||:|  ||:|| ||:|   | |||:|||||  : || |     :
Db     562 LDFSRRADALKLWDLSPLQYVENVKTPTLIVHSVLDHRCPVEQAEQWYAALHKHQVPVRF 621

Qy     624 VRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
           ||||  :| |||:| | :|: |||   :| :: |
Db     622 VRFPEENHELSRSGRPDRRLTRLNEYFAWLERWL 655


RESULT   4
Q9YAW3
ID   Q9YAW3      PRELIMINARY;      PRT;   659 AA.
AC   Q9YAW3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   659AA LONG HYPOTHETICAL ACYLAMINO-ACID-RELEASING ENZYME.
GN   APE1832.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000062; BAA80835.1; -.
DR   InterPro; IPR000379; -.
SQ   SEQUENCE   659 AA;  73895 MW;  475E78DEE5863551 CRC64;


  Query Match             27.6%;  Score 962.5;  DB 1;  Length 659;
  Best Local Similarity   34.8%;  Pred. No. 2.3e-58;
  Matches  235;  Conservative  120;  Mismatches  274;  Indels   47;  Gaps   18;

Qy       2 KKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWI-----YETKTGGS 56
           || :| :||  :  |::|  :|:| : ||: :: ::::::| | ||:     |   |||
Db       5 KKPVTPEDIIRLTFVSNPSVSPEGDKVAYLATKADEKENTYRSGIWLAEEDSYRPLTGG- 63

Qy      57 VPWTHGEKRSTDPRWSPDGRTLAF---ISDREGDAAQLYIMSTEGGEARKLTDIPYGVSK 113
            |:         | |||||  :||   ::  :     : |:| :||| : | :     :
Db      64 -PY------DRCPAWSPDGEIIAFSRTVNSNQRKRHYIAIVSPKGGEPQVLHESVSPATS 116

Qy     114 PLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVL 173
             ||  |  |         |     ||        : :|:  :    | :|     |  | :
Db     117 IKWSSQGSMIGYLSRKPTG-----REWKPYSERDVLEIDRIPVWFDSEGWVFDRYWGLTV 171

Qy     174 VSVKSGE-MKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKP--HDVYIMSLESGD 230
           :|   || : |       :  |  |:||  :|::     :  |  ||  | : :  | ||
Db     172 ISYPGGEVLLEKGGVNYNIVDFDFAPDDNTIVYAV----STDMKKPFIHKLVLWDLASG- 226

Qy     231 LKQVTPHRG-SFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDV 289
           :: |   | :  : :| | |||:|:  |:::     :  | ::||:      |||  ||:
Db     227 -REKTLIEGLTIAAVAFDPRGRYIAVKANDRKRGLFSHYKIFVYDLNSEEFVCLTCDLDL 285

Qy     290 HLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYIS-----IEGLVYPIRLE 344
           :  : :  |:       :   |  |  |    |       :   |       | :: |
Db     286 NTLNTVNSDARGPSCLRGMYW--DDNGHLYYGVHNAGRMVVMKSRPLGEAEAVLDP---S 340

Qy     345 KEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEI 404
           :: ||:|             |  | ::| |     | :||  |  |    ::: |  :
Db     341 SATVDDFSISRGGDTIAYVKMGPTSPPDIY-IYRDDNEYRLTDHNAWFAESRSLAEPVRL 399

Qy     405 QYATEDGVMVNGWLMRPAQMEGETT--YPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAV 462
           |  :  |  :: |:: |    || :   | || |||||   ||: : ||||:||::|:||
Db     400 QVQSPLGGSIDAWILLPPD-AGECSGCLPWILYIHGGPKTSYGYAFIHEFQMLASQGFAV 458

Qy     463 VYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGF 522
           :| ||||| || :|| : :|| ||  || ::|: ||||:   | :||:| ||||||||:
Db     459 IYSNPRGSDGYSEEFAD-LRGRYGVDDYSELMKVVDEALSDFPQLDPQRGGVTGGSYGGY 517

Qy     523 MTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE-HDMFEDTEKLWDRSPL 581
           ||| || :| |||||||||| ||||||:| ||||::|    :   : : | ||  | |||
Db     518 MTNVIVTKTRRFKAAVTQRSCSNWISFYGESDIGWYFAPELISAQEPWRDLEKYVDFSPL 577

Qy     582 KYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQ 641
             |:||||||:|   | |||:  | ||| |||  | ||:|: ||  :|||||:| |::
Db     578 FSVENIETPLLIIHSTEDFRCPLSGAIQLFTALKLKGVETRLLVFPGENHNLSRSGTPKR 637

Qy     642 RIKRLNYISSWFDQHL 657
           |: ||  |:||| :||
Db     638 RVARLRAIASWFKEHL 653


RESULT   5
Q9HN07
ID   Q9HN07      PRELIMINARY;      PRT;   674 AA.
AC   Q9HN07;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
```

```
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ACYLAMINOACYL-PEPTIDASE.
GN   YUXL OR VNG2302G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC   Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005114; AAG20414.1; -.
DR   InterPro; IPR000379; -.
SQ   SEQUENCE   674 AA;  72577 MW;  DCBEEF48896CCA78 CRC64;

 Query Match             25.9%;  Score 903;  DB 1;  Length 674;
 Best Local Similarity   31.7%;  Pred. No. 3.1e-54;
 Matches  223;  Conservative  121;  Mismatches  251;  Indels  108;  Gaps   22;

Qy       5 ITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDS--YTSNIWIYETKTGGSVPWTHG 62
Db       4 VEASDYWELARLSQPAVSPTGDRVAYVRTTAERDGDTGQETASVHVAPVGGGGSRRFTAA 63

Qy      63 EKRSTDPRWSPDGRTLAFISDREGDAA----QLYIMSTEGGEARKLTDIPYGVSKPLWSP 118
Db      64 DGTDAEPAWSPTGDRLAFVSTR-GDAGGGRPQLWVVPVDGGEATRITDVIGGVSNIAWGP 122

Qy     119 DGESILVTISLGEGESID--DRE---KTEQDSYEPVEVQGLSYK-----RDGKGLTRGAY 168
Db     123 DGDRVAFLQAVRPAERAEELDREHDGEYERDTPDPRVIDRQMYRGHTSYRD------GAR 176

Qy     169 AQLVLVSVKSGEMKELTSHKADHGDPAFSPDG----------------KWLVFSANLTET 212
Db     177 RHVYVADLAAGRVDRVTAGVAECGGPAWGDDGTLYYPIRRGLDADDRLEWAI-------- 228

Qy     213 DDASKPHDVYIMSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNA------- 265
Db     229 -EAATPAD-------SADSTVVTTVEGM---------GPTLAVHGSRVAYTTTPADDPTL 271

Qy     266 TLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQG 325
Db     272 TQTAVEVFDRDTGATDRLTDGVDRTVAERTAGHA--------PEWGPDGEHLYVCTPDEG 323

Qy     326 ---------STGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSELY-S 375
Db     324 GYVLRRVAADTGDDRVVLAG--------DRHVHAFSVSADAVAVVQS--EWDHPGEVVAA 373

Qy     376 IPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVM--VNGWLMRPAQMEGETTYPLI 433
Db     374 TPMGGEEVTLTRANRDYLDERAVSKPESVWI---DGPAGDIQGWVLTPPEFDPEETYPLV 430

Qy     434 LNIHGGPHMMY--GHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYD 491
Db     431 VNVHGGPHLMWSAAGTMWHEFQTLAARGYVVFWCNPRGSTGYGQEWLRAVERDWGAVTLA 490

Qy     492 DVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHG 551
Db     491 DVMAGVAAVTDRD-YVDADNAFITGGSFGGFMTGWAVGHTDAFAGAVAQRGVYDLTGFYG 549

Qy     552 VSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLF 611
Db     550 TTD-AYALVEGEFDTDPVRDNAFLWAQSPAAHTDAVDTPTLLVHSEADYRTPANTAELFY 608

Qy     612 IALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFD 654
Db     609 RLLKKHGVDTRLVRYPGEGHDLSRSGDPGHVVDRIERIAAWFD 651

RESULT   6
Q9Y9J8
ID   Q9Y9J8      PRELIMINARY;      PRT;   572 AA.
AC   Q9Y9J8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   572AA LONG HYPOTHETICAL ACYLAMINO-ACID-RELEASING ENZYME.
GN   APE2290.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000064; BAA81302.1; -.
DR   InterPro; IPR000379; -.
SQ   SEQUENCE   572 AA;  62902 MW;  1EEE3752567B9510 CRC64;

  Query Match             20.0%;  Score 697;  DB 1;  Length 572;
  Best Local Similarity   33.5%;  Pred. No. 4.4e-40;
  Matches  199;  Conservative   99;  Mismatches  248;  Indels   48;  Gaps   20;

Qy      69 PRWSPDGRTLAFISDR---EGD-AAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESIL 124
Db       2 PSWSPDASKLAFASRRGSGKGEKGSGLYIVG-RGGEPRRVAWFTHGVSELGWA-SATKVY 59

Qy     125 VTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKEL 184
Db      60 VVAPVAD-------EKTRDKDGDYVATDRLPLWFDSSGLVAGLADQIHLVDAESGRTSRL 112

Qy     185 TSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGSFGSS 244
Db     113 TSE--DLGVLTAEPCGN-SIYYVTLRRWAD---PLDTVVKSVPISGGEPETVLEGLTVSQ 166

Qy     245 SFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLT-EMLDVHLADALIGDSLIGG 303
Db     167 LRCINGK-LFMLAHKREIGLASHSKLYLLE-DRGKASCLTCGILDRNIW------SIAGG 218

Qy     304 AEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIAS 363
Db     219 LGGEPV---------IVYADRGRSVI--AAVEGRVEDLVRGDMIVVQADSNGEEVVYIAS 267

Qy     364 VTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQ 423
Db     268 --SPTEPPEVYRFKGGDVER-VSSINRWVVEEFRLYKPRRVEVEAE-GEVVEGWYIEP-- 321

Qy     424 MEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRG 483
Db     322 -EGEGRKPLILFIHGGPKGMYGYGFYGEMQLFASEGFTVAYANPRGSDGYEEEFAD-IRG 379

Qy     484 DYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSI 543
Db     380 RYGDTDYKQLMKFLD-AVVEGSDVDEGKMAVTGISYGGYMTNVIVTKTSRFRAAVSENGI 438

Qy     544 SNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCP 603
```

```
          ::||:    :||||:|   |:     :: |:   |||  :| || ||::::|   | ||
Db    439 ADWIADFWAADIGYWFDPDQIGGTPLDNLEEYVKRSPAFHAKNVNTPVMLIHSMEDYRCF 498

Qy    604 IEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
          |:||  :  |||  |||: || |   ||  |    || | |||     |  : |
Db    499 IDQALAMHTALKVNGKESFLVVFTKGSHGHSILASPRHRRKRLELKLKWIKEKL 552


RESULT   7
Q9UYQ8
ID   Q9UYQ8      PRELIMINARY;      PRT;   632 AA.
AC   Q9UYQ8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   PEPTIDASE.
GN   PAB1418.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORSAY;
RA   Heilig R.;
RT   "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT   structure and evolution.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ248287; CAB50354.1; -.
DR   InterPro; IPR000379; -.
DR   InterPro; IPR000834; -.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
SQ   SEQUENCE   632 AA;  74117 MW;  BEBB61002D6BE55B CRC64;


  Query Match             19.1%;  Score 667.5;  DB 1;  Length 632;
  Best Local Similarity   28.4%;  Pred. No. 5.7e-38;
  Matches  194;  Conservative  132;  Mismatches  263;  Indels   93;  Gaps   27;

Qy      1 MKKLITADDI----------TAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY 49
          | :||  | :          |    ::||:    |:  ||| :: | : : | | | :
Db      1 MMQLIFGDTVMTSIEWNEKTFTKFAYLSDPRIR--GSTIAYVLTKANLDNNKYESTI-VL 57

Qy     50 ETKTGGSVPWTHGEKR----STDPRWSPDGRTLAFIS-DREGDAAQLYIMSTEGGEARKL 104
          |    ||       ||    :: || ||||: :||:  : |   ||:::   :   |:|:
Db     58 ENLEDGS-------KRFIEDASMPRISPDGKKIAFMRFNEEKKVAQIWVADMKTLSAKKV 110

Qy    105 TDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLT 164
           :    :    |: |   :||          |  : : ::|     :|        |  |
Db    111 LEAK-NIRSLEWNDDSRRLLV---------IGFKRRDDEDFIFEDDVPAWF---DNMGFF 157

Qy    165 RGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKW----LVFSANLTET-DDASKPH 219
           |    :|  :: |: |          | ||  | |    :| :    :|      |
Db    158 DGEKTTFWIVDTEAEEVIE------QFEKPRFS-SGVWHGDSIVVNVPYRDTIPRYFKYW 210

Qy    220 DVYIMSLESGDLKQVTPHRGSFGSSSF---SPDGRYLALLGN-EKEYKNATLSKAWLYDI 275
          ::|:   : |: :::      |   ||     ||: : | |  ||:| :    | ::||
Db    211 NIYLW--KDGEEEKL------FEKVSFHAIDSDGKNILLYGKPEKKYMSEH-DKLYIYD- 260

Qy    276 EQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIE 335
           | :  : : :|     | |               || : :: :  ::||  :|    :
Db    261 --GEVKGIMDSIDREAGQAKI--------------KDGKVYFTL-YEEGSVNLYL--WD 300

Qy    336 GLVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVRE 395
          | :  |   | :|  | :    :   : |   | |||      ||:|||  |     :
Db    301 GDIKEIAKGKHWIMGFDVDEKVVYLKETAT---RLRELYI--WDGEERQLTDYNGLIFSK 355

Qy    396 HTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVL 455
           |:  :: : | : :|| ::|   |||   |:|: :||||  |||: : :| |::
Db    356 LKTFEPKHFKFKSID-LEIDGWYIKPELKEGEKA-PVIVFVHGGPKGMYGYYFKYEMQLM 413

Qy    456 AAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVT 515
          |:||| :||:|||||:|| ::|   |    | :|: |::  ::|  | :|  | :|:|:|
Db    414 ASKGYYIVYVNPRGSNGYSEDFALRVLERTGLEDFQDILNGIEEFFKLEPQADRERVGIT 473

Qy    516 GGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKL 575
          | |||||||| : |:: ||| :::  || |:: :  |||| :|    :  :  |: |
Db    474 GISYGGFMTNWALTQSDLFKAGISENGISYWLTSYAFSDIGLWFDKEVIGENPLEN-ENF 532

Qy    576 WDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSR 635
           ||| || ||: |||::|   | |||::|:   :  || :|||  :  |   :|  |
Db    533 RKLSPLFYAKNVKAPLLLIHSLEDYRCPLDQSLMFYHVLKDLGKEVYIAIFKRGAHGHSI 592

Qy    636 TGHPRQRIKRLNYISSWFDQHL 657
           | || |:||      :|:: |
Db    593 RGSPRHRMKRYKLFMEFFERKL 614


RESULT   8
O58323
ID   O58323         PRELIMINARY;   PRT;   622 AA.
AC   O58323;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 73.2 KDA PROTEIN PH0594.
GN   PH0594.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000002; BAA29683.1; -.
DR   InterPro; IPR000379; -.
KW   Hypothetical protein.
SQ   SEQUENCE   622 AA;  73231 MW;  6D4BAB1A12614E02 CRC64;


  Query Match             18.9%;  Score 660;  DB 1;  Length 622;
  Best Local Similarity   28.2%;  Pred. No. 1.8e-37;
  Matches  189;  Conservative  130;  Mismatches  251;  Indels  100;  Gaps   26;

Qy     11 TAIVSVTDPQYAPDGTR---AAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRST 67
          |    ::||:     ||    ||| :: | | : | : | : |    ||  :    : ::
Db     12 TKFAYLSDPR-----TRKNLVAYVLTKANLESNKYENTI-VIENLEDGSRKFI---EDAS 62

Qy     68 DPRWSPDGRTLAFIS-DREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVT 126
           || ||||: :||:  : |   ||:::   :   |:|: :    :    |: |   :|
Db     63 MPRISPDGKKIAFMRFNEEKKTAQIWVADLKTLSAKKVLEAK-NIRSIEWNQDSRRLL-- 119

Qy    127 ISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTS 186
                ::  : : ::|     :|       | |   |      ::  :  |: |
Db    120 -------AVGFKRREDEDFIFEDDVPAWF---DNMGFFDGEKTTFWVIDTEGEEVIE--- 166

Qy    187 HKADHGDPAFSPDGKW----LVFSANLTETDDASKPH----------DVYIMSLESGDL 231
                | ||  | |    :| |           ||           |:|:   : |:
Db    167 ---QFEKPRFS-SGIWHGDSIVVSV----------PHRDVIPRYFKYWDIYLW--KDGEE 210

Qy    232 KQVTPHRGSFGSSSF---SPDGRYLALLGN-EKEYKNATLSKAWLYDIEQGRLTCLTEML 287
          :::      |   ||     ||  : | |  ||:| :    | ::||   | :  : : :
Db    211 EKL------FEKVSFYAIDSDGERILLYGKPEKKYVSEH-DKIYIYD---GEVKGILDDI 260

Qy    288 DVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEY 347
          |  :| | | :  :               |    ::||  :|    :| |  |   | :
Db    261 DREVAQAKIRNGKV---------------YFTLFEEGSVNLYL--WDGEVREIAKGKHW 302
```

```
Qy    348 INSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYA 407
          |  |      ::  |       ||:|||      ||:|||  |    ::     |   ::
Db    303 IMGFDA---DERLIYLKETATRPAELYL--WDGEERQLTDYNGLIFKKLKTFEPRHFRFK 357

Qy    408 TEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINP 467
          : | : ::|| ::|   |||   |:|: :|||| |||: : :| |::|:||| :||:||
Db    358 SID-LELDGWYIKPEIKEGEKA-PVIVFVHGGPKGMYGYYFKYEMQLMASKGYYIVYVNP 415

Qy    468 RGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWI 527
          |||:|| ::|   |    | :|: |::  ::| :: :|  | :|:|:|| ||||:||||
Db    416 RGSNGYSEDFALRVLERTGLEDFQDILNGIEEFLRLEPQADRERIGITGISYGGYMTNWA 475

Qy    528 VGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANV 587
          : |:: ||| :::  || |:: :  |||| :| | ::  |   : |     ||| || ||
Db    476 LTQSDLFKAGISENGISYWLTSYAFSDIGLWF-DKEVIGDNPLENENYRKLSPLFYAKNV 534

Qy    588 ETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLN 647
          : |||::|   | |||::|:   :  || :|||  :  |   :|  |  | || |:||
Db    535 KAPLLLIHSLEDYRCPLDQSLMFYHVLKDLGKEVYIAIFKKGAHGHSIRGSPRHRMKRYK 594

Qy    648 YISSWFDQHL 657
            :|:: |
Db    595 LFMEFFERKL 604
```

RESULT 9
Q9PB84
ID Q9PB84 PRELIMINARY; PRT; 709 AA.
AC Q9PB84;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ALANYL DIPEPTIDYL PEPTIDASE.
GN XF2260.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE004038; AAP85059.1; -.
DR InterPro; IPR000379; -.
SQ SEQUENCE 709 AA; 78678 MW; 99CF7D35B0F96A75 CRC64;

Query Match 13.6%; Score 473; DB 2; Length 709;
Best Local Similarity 22.7%; Pred. No. 2e-24;
Matches 163; Conservative 138; Mismatches 288; Indels 130; Gaps 23;

```
Qy      9 DITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGS--------VPWT 60
          |: |:  |: |: :|||    : | |:: :    ::::|:   : | |        : |
Db     49 DMVALDRVSSPELSPDGAVLVFAKRQMDAKYIKASTSVWVQRLQAGTSAAPVRLTPLGWD 108

Qy     61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
              : | :| ||: : |:| : | : |||::   || :|:|||::   :     || |
Db    109 -----VSAPVFSRDGKAVYFLSAKSG-SHQLYVLPVSGGTSRQLTNLAVDIDSYKLSPQG 162

Qy    121 ESILVTISL----GEGESIDDREKTEQDS-------YEPVEVQGLSYKRDGKGLTRGAYA 169
          : |: :  :     |   |   |:  |: :       :| : |:      ||:  |    |
Db    163 DRIVFSAGVFQVCGSDLSCTKRKLDEKKNAKASGVVFEQLFVRHWDTWNDGRRNTL-FIA 221

Qy    170 QLVLVSVK------------SGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDA-S 216
          |  |  |               |::        |||    :|||| :| |  :    :  |
Db    222 SLPAVGAKPVSVVSAMSAMLDGDVPSKPFGGADHF--VWSPDGHSVVASIRVAGRQEPWS 279

Qy    217 KPHDVYIMSLESGD----LKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWL 272
             |:|   :   |    |    |   :  :  || ||: |      :    |
Db    280 TNFDLYRFDVSGHDAPVNLTAANPAWDA--TPMFSADGKMLYYRAMRRPGFEADRFGLME 337

Qy    273 YDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTK---------DSQGFYVIGTD 323
          ::: |::                          |  | | :         | :  ||   |
Db    338 MEVQSGKV----------------------REIAPHWDRSADEIALSADGKALYVNADD 374

Qy    324 QGSTGIYYISI----------EGLVYPIRLE--KEYINSFSLSPDEQHFIASVTKPDRPS 371
          |   :: : |           || |:   |   |      ||   :| |:
Db    375 HGEHPLFKVDIASGKVEKWVGEGSVHAPVLAGGKLAFARNSLKSADQIFVTDAVARG--- 431

Qy    372 ELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYP 431
            ||    : :| |  : :::  :   |:  :   :   | ||:::|   :    ||
Db    432 -----PL----QAITSATGEVLQQVRLGDFEQFSFKGWNDETVYGYVVKPYDYQPGKKYP 482

Qy    432 LILNIHGGPHMMYGHTYFHEF--QVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKD 489
          :   |||||   :|::: : :  |  | :|||||  |:  || |||| | :|:   :| :
Db    483 VAFLIHGGPQGSFGNSWGYRWNPQTYAGQGYAVVMIDFHGSTGYGQAFTDAISQHWGDRP 542

Qy    490 YDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNW--- 546
            :|: :    | :: | ::  :    | ||||:|  || |  |:            |
Db    543 LEDLQKGWAAAQQQYPFLNGDKACALGASYGGYMVYWIAGHWNQ----------PWKCL 591

Qy    547 ISFHGVSD---IGY-----FFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
          :   || |   :||     :|::|:     :|:       :|: :      |:|::||::
Db    592 VDHDGVFDNRMMGYATEELWFSEWENGGTPWENPAGYEQFNPVLHVDKWRVPMLVIHGQK 651

Qy    599 DDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
          | | |:||    | ||:: | |:||: | : :| : :   |:  |:  : :::|  :::
Db    652 DFRIPVEQGLAAFGALQRKGIESKLLYFHDENHWVLK---PQNSIQWHDTVNAWLKKYI 707
```

RESULT 10
Q9U592
ID Q9U592 PRELIMINARY; PRT; 663 AA.
AC Q9U592;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DIPEPTIDYL-PEPTIDASE (FRAGMENT).
GN DPP.
OS Entamoeba dispar.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=46681;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAW142;
RA Bruchhaus I., Hellberg A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF059279; AAF20265.1; -.

```
DR   InterPro; IPR000379; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002470; -.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PRINTS; PR00862; PROLIGOPTASE.
FT   NON_TER       1       1
SQ   SEQUENCE   663 AA;  76572 MW;  25EF2B47D0CC36DF CRC64;


  Query Match             12.6%;  Score 439;  DB 5;  Length 663;
  Best Local Similarity   24.5%;  Pred. No. 4e-22;
  Matches  167;  Conservative  119;  Mismatches  270;  Indels  126;  Gaps   28;

Qy      42 YTSNIWI-YETKTGGSVPWTHGEKRSTDPRWSP-------DGRTLAFISDREGDAAQLYI 93
           |||: |   | ||       |  |: |      |      :     |:   ||    :|:
Db      39 YTSSQWFEEENKT------KHSLKKLTLTETEPIDLVPYSEESFSNFVCLNEG----VYL 88

Qy      94 M------STEGGEARKLTDIPYGVSK-PLWSPDGE-SILVTISLGEGESIDDREKTEQDS 145
           :      |  :   :|  |  :    : | ||:   | ::::  | ::::  |
Db      89 LHNGQIAKIENDKIIDITSTPISIDTFEMISVDGKLRGLASMTVFPGMTLEESAK----- 143

Qy     146 YEPVEVQGLSYKRDGKGLTRGAYAQLV--------------LVSVKSGEMKELT----SH 187
           : :|::  :|:          | ||:              | : |  | |  |     :
Db     144 -KFIELEKQNYR---------VYDQLMIRRWDTYWDGQFQHLFTFKQEEDKSFTFKDIMN 193

Qy     188 KADHGDPA----------FSPDGKWLVFSANLTETDDASKPHDVYIMSLESG-DLKQVT- 235
           :  ||           ||||  : ||  | |   :|  ::||  : ::  :   :|
Db     194 DKKYDCPARPFGGNEEYDISPDGTKIAFSI-LLENPASSLDNNVYYATFDNPLEWNIITI 252

Qy     236 PHRGSFGSSSFSPDGR---YLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLD---- 288
           ::|      |: ||    ||::   : |   : |      | ::  :  :|  :|
Db     253 DNKGYDNQPLFNNDGSLLYYLSMSAPKDESDKSVLKSX---DFKEKIIRDITGKIDLSFS 309

Qy     289 ---------VHLADALIGDSLIGGAE-QRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLV 338
                   :||:  : |:  | | :   :   |::      | | :|: | :  :  |::
Db     310 APMKIVGNYMHLSTQIEGNVYIVGLDTSKSELTREDVN---IITQKGTAGSFVFTTTGII 366

Qy     339 YPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTI 398
           |      || |||:|                  | ||: |   :  ||:|  | : :
Db     367 Y------EY-NSFTL----------------PQELF-IFENKVIKQITHINQEVLSTIKF 402

Qy     399 SIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEF--QVLA 456
           :|: |   :  :: ::  |  |:  | ||:|| ||||   : : : : :  ||:|
Db     403 GEYKEVHYTGANNDQIHAFITYPPNMKTNTKYPVILYTHGGPESPWTNNFHYRWNPQVIA 462

Qy     457 AKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTG 516
           |:|| |   |  ||  || ||: |:| ::||  ::|:|: :|     :| :|       |
Db     463 AQGYIVFAPNFHGSGSYGDEFLKAIRKNWGGWPFEDLMKGMDYLKTNEPIVDIDNACAMG 522

Qy     517 GSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKL 575
           ||||:| |||  | | |||  :   | :   :  |  ||  : :  : |:||
Db     523 ASYGGYMMNWINSQNTGRFKCIICHDGIMDSEGSYYYMDEMYFL-ETEFGYPMYEDNTYY 581

Qy     576 WDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSR 635
           |||  |     || | :||  | |  :    | |: |::   |:||| :|  :| : |
Db     582 KKYSPLNYVNGYNTPQLTIHGGSDYRIDVVAGYQQFVTLQRKNIESKLVIYPEENHWVLR 641

Qy     636 TGHPRQRIKRLNYISSWFDQHL 657
           |   |     : :|  ::|
Db     642 ---PYNSIDWHTQVFNWLAKYL 660


RESULT  11
Q9U593
ID   Q9U593      PRELIMINARY;      PRT;   665 AA.
AC   Q9U593;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   DIPEPTIDYL-PEPTIDASE.
GN   DPP.
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=5759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM-1:IMSS;
RA   Bruchhaus I., Hellberg A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF059278; AAF20264.1; -.
DR   InterPro; IPR000379; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002470; -.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PRINTS; PR00862; PROLIGOPTASE.
SQ   SEQUENCE   665 AA;  76529 MW;  D5D71BD45D42AA94 CRC64;


  Query Match             12.4%;  Score 433.5;  DB 5;  Length 665;
  Best Local Similarity   27.7%;  Pred. No. 9.7e-22;
  Matches  134;  Conservative   84;  Mismatches  207;  Indels   59;  Gaps   16;

Qy     197 SPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQ---VTPHRGSFGSSSFSPDGR-- 251
           ||||  : ||  | |   :|  ::||  : |:  |:     | ::|      |: ||
Db     215 SPDGTMIAFSI-LLENPASSLDNNVYYATFEN-PLEWNCITTNNKGYDNQPLFNNDGSLL 272

Qy     252 -YLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLD------------VHLADALIG 297
            ||::   : |   : |      ||  |  :  :|  :|             :||:  : |
Db     273 YYLSMSAPKDESDKSVLKS---YDFNQKVIRDITGNIDLSFSAPMKVVGNYMHLSTQIEG 329

Qy     298 DSLIGGAE-QRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPD 356
           :  | | :  :   |::      | | :|: | :  :  |::|      || |||:|
Db     330 NVYIVGLDTSKSELTREDVN---IITQKGTAGSFVFTTTGIIY------EY-NSFTL--- 376

Qy     357 EQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNG 416
                     | ||: :   :  ||:|  | : :         :|| |   :   ::
Db     377 ------------PQELF-LYENKVIKQITHINQEVLSTIKFGEYKEIHYTGANNDQIHA 422

Qy     417 WLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEF--QVLAAKGYAVVYINPRGSHGYG 474
           ::  |  |:  | ||:||  ||||   : : : : :  ||:||:|| |   |  ||  ||
Db     423 FITYPPNMKTNTKYPVILYTHGGPESPWTNNFHYRWNPQVIAAQGYIVFAPNFHGSGSYG 482

Qy     475 QEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQ-TNR 533
             |: |:| ::||  ::|:|: :|     :| :|       | ||||:| |||  | | |
Db     483 DAFLKAIRKNWGGWPFEDLMKGMDYLKTSEPLVDIDNACAMGASYGGYMMNWINSQNTGR 542

Qy     534 FKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLI 593
           ||  :   | :   :  |  ||  : :  : |:||       ||| |     || |
Db     543 FKCIICHDGIMDSEGSYYYMDEMYFL-ETEFGYPMYEDDTYYKKYSPLNYVNGYNTPQLT 601

Qy     594 LHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWF 653
           :|| | | :    | |: |::   |:||| :|  :| : |   |  |  :  |
Db     602 IHGGSDYRIDVVAGYQQFVTLQRKNIESKLVIYPEENHWVLR---PYNSIDWHAQVFDWL 658

Qy     654 DQHL 657
           ::|
Db     659 AKYL 662


RESULT  12
Q9Y944
ID   Q9Y944      PRELIMINARY;      PRT;   591 AA.
AC   Q9Y944;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   591AA LONG HYPOTHETICAL ACYLAMINO-ACID-RELEASING ENZYME.
GN   APE2441.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
```

```
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000064; BAA81456.1; -.
DR   InterPro; IPR000379; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002925; -.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   Pfam; PF01738; DLH; 2.
SQ   SEQUENCE   591 AA;  65756 MW;  ECA77CA74A401D67 CRC64;


 Query Match             11.8%;  Score 410;  DB 1;  Length 591;
 Best Local Similarity   24.9%;  Pred. No. 3.4e-20;
 Matches  167;  Conservative   99;  Mismatches  244;  Indels  162;  Gaps   26;

Qy      12 AIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRW 71
           |: |   |:  |  |   |  |      || :|   |:::    :  |   |:|
Db       8 AVKSSHTPRIDP-STGCIYYLS------DSASSQPVIWKSCESRNDVWLPWERRVGSLEI 60

Qy      72 SPDGRTLAFISDREGDAA-QLYIMSTE------GGEARKLTDI-PYGVSKPLWSPDGESI 123
           :|||  :|| :|::||    :||   :       ||   :  | |       |||||  :
Db      61 APDG-LVAFSTDKDGDERWSIYIAGEDGSINMIGP-------WSPDGRLL 112

Qy     124 LVTISLGEGESID----DREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSG 179
             | :   |   |    ||||            :|   :|:|:   :
Db     113 AFTSNKRNGVDFDLYVFDREK-----------GSVSIVVEGEGIIAAS------------ 149

Qy     180 EMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRG 239
                             |:  |  |    :::   |:  ::: |: | :| |||
Db     150 ----------------------SWIDGSRLLAVKRNSNLDSDILAVNVYKGEAKVLTRHRG 188

Qy     240 SFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDS 299
              ::|  | |   ||  :  :                        :|   : | |   ||
Db     189 EELNTSPRPIGGGKALFISNMD----------------------SEFTGIALIDLQTGDR 226

Qy     300 LIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEG--LVYPIRLEKEYI--------- 348
           :  ||   |  | :   |     | | :| ::::|   ||  ::::
Db     227 RLVFREQ---W--DVEALEV----SGKTVVYSLNVDGESRVYTTKIDERGSLLQPRPIEG 277

Qy     349 ----NSFSLSPDEQHFIA--SVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPE 402
               :  ||  :|:  :|  |:: |    |:|   ||:   :|| :   ::||     ||
Db     278 LPRGSLLSLDLNEKLGLAVFSLSTPKHGIEIYIAGLGRSASRLTVSPKAWLREDEFVEPE 337

Qy     403 EIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEF-QVLAAKGYA 461
             | : ||: |   | :|      |  | :: :||||        |: | | ||| | |
Db     338 HFSYESFDGIKVRALLYKP-DKPLYTPPPAVVYLHGGPESQ-ERVRFNVFPQALAAIGIA 395

Qy     462 VVYINPRGSHGYGQEFV---------NAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRL 512
             :  | ||| |||: ||          :|||          ||  ||  |::    :|  ||
Db     396 TIAPNYRGSTGYGRRFVHLDDVEKRMDAVR---------DVYYAVKAAVEAG-LVDGSRL 445

Qy     513 GVTGGSYGGFMTNWIVG-QTNRFKAAVTQRSISNWISF----------HGVSDIGYFFTD 561
             | ||||||::|   :    : :|  |    | | ::|          : ::: |
Db     446 CVMGGSYGGYLTLMSLAIYPDLWKCGVEIVGIVNLVTFIRNTSPYRRRYRIAEYG----- 500

Qy     562 WQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKET 621
                 |    | :   ||: |  |:: ||:::|| :| | |: :|||  ||   |
Db     501 -----DPDVHGEIMLKLSPITYVENMKAPLMVIHGAKDPRVPVSEAEQLVEALSSRGVRV 555

Qy     622 KLVRFPNASHNL 633
           : ||  :  | :
Db     556 RYVRLEDEGHGI 567


RESULT  13
O13479
ID   O13479      PRELIMINARY;      PRT;   721 AA.
AC   O13479;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   DIPEPTIDYL-PEPTIDASE V PRECURSOR (EC 3.4.14.-) (DPP V).
OS   Aspergillus fumigatus (Sartorya fumigata).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5085;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 19-42; 435-448 AND 593-603.
RC   STRAIN=CBS 144.89;
RX   MEDLINE=97197790; PubMed=9045640;
RA   Beauvais A., Monod M., Debeaupuis J.-P., Diaquin M., Kobayashi H.,
RA   Latge J.-P.;
RT   "Biochemical and antigenic characterization of a new dipeptidyl-
RT   peptidase isolated from Aspergillus fumigatus.";
RL   J. Biol. Chem. 272:6238-6244(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN METABOLISM OF DIPEPTIDES OR MAY
CC       AFFECT HOST DEFENSE MECHANISMS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MYCELIA AND CONIDIA.
CC   -!- PTM: GLYCOSYLATED.
CC   -!- MISCELLANEOUS: THE OPTIMUM PH IS 6.5.
DR   EMBL; L48074; AAB67282.1; -.
DR   MEROPS; S09.012; -.
DR   InterPro; IPR000379; -.
KW   Hydrolase; Signal; Antigen; Glycoprotein.
FT   SIGNAL       1     18
FT   CHAIN       19    721       DIPEPTIDYL-PEPTIDASE V.
FT   ACT_SITE    560    560       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    643    643       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    675    675       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     75     75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     94     94       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    380    380       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    450    450       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   721 AA;  79657 MW;  93E001B51111896E CRC64;


 Query Match             11.6%;  Score 405.5;  DB 3;  Length 721;
 Best Local Similarity   22.3%;  Pred. No. 9.6e-20;
 Matches  161;  Conservative  107;  Mismatches  298;  Indels  155;  Gaps   26;

Qy       6 TADDITAIVSVTDPQYA-----PDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWT 60
             ||  :|    :| |: :     | |  | :  || : |    ||   : : |||  :   |
Db      15 TALALTPEQLITAPRRSEAIPDPSGKVAVFSTSQYSFETHKRTSWWSLLDLKTGQTKVLT 74

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGD---AAQLYI------------------------ 93
           : :   ::  |  |  :: :::    |     :|::
Db      75 N-DSSVSEIVWLSDD-SILYVNSTNADIPGGVELWVTQASSFAKGYKAASLPASFSGLKA 132

Qy      94 MSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQG 153
             |: |: |    : || | |  :   |  | | |  |       |       | :
Db     133 AKTKSGDIR---FVAYGQSYPNGTAYNEE-LATAPLSSARIYDSIYVRHWDYWLSTTFNA 188

Qy     154 LSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTS--HKADHGDPAF--------SPDGKWL 203
           :      ||  :  |:          ||:| | |    |:   | |        ||||||:
Db     189 VFSGTLKKGHGKNGYS-------LDGELKNLVSPVKNAESPYPPFGGASDYDLSPDGKWV 241

Qy     204 VFSANLTETDDAS--------KPHDVYIMSLE-SGDLKQVTPH--RGSFGSSSFSPDGRY 252
            | :   |   |:          |||    :   :|     ||   :|   |  |||:|
Db     242 AFKSKAPELPKANFTTSYIYLVPHDASETARPINGPDSPGTPKGIKGDSSSPVFSPNGDK 301

Qy     253 LALLG-NEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRP--- 308
           ||      ::  |::    :| ||    |    :                 |: |  :: |
Db     302 LAYFQMRDETYES---DRALLYVYSLGSKKTI-------------PSVAGDWDRSPDSV 344

Qy     309 IWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEY----------INSFSLSPDEQ 358
```

```
          || | :   |   | | | :: :       |   : :|          ::::   ||
Db    345 KWTPDGKTLIVGSEDLGRTRLFSL-------PANAKDDYKPKNFTDGGSVSAYYFLPDSS 397

Qy    359 HFIAS----------VTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYAT 408
            :              ||::          : : :| |     : |           |
Db    398 LLVTGSALWTNWNVYTAKPEKGVIKKIASANEIDPELKGLGPSDISEF---------YFQ 448

Qy    409 EDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEF--QVLAAKGYAVVYIN 466
          :   :: |:: |   :    ||||  |||||   :   :   :  | :|| ||  |
Db    449 GNFTDIHAWVIYPENFDKSKKYPLIFFIHGGPQGNWADGWSTRWNPKAWADQGYVVVAPN 508

Qy    467 PRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNW 526
          | || |:||    |:: ::||  |||:::  :   :   ::|       | ||||||  ||
Db    509 PTGSTGFGQALTTAIQNNWGGAPYDDLVKCWEYVHENLDYVDTDHGVAAGASYGGFMINW 568

Qy    527 IVGQT--NRFKAAVTQRSISNWISFHGVSDIGYFFTD--------WQLEHDM------FE 570
          | |     :|||                ||  | |  |         | :: :
Db    569 IQGSPLGRKFKAL-------------VSHDGTFVADAKVSTEELWFMQREFNGTFWDAR 614

Qy    571 DTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNAS 630
          |  : || |  :      ||:|::| ::| | |: :   ||  |:: |  :: : ||: :
Db    615 DNYRRWDPSAPERILQFATPMLVIHSDKDYRLPVAEGLSLFNVLQERGVPSRFLNFPDEN 674

Qy    631 H 631
          |
Db    675 H 675


RESULT  14
Q9UW98
ID   Q9UW98      PRELIMINARY;      PRT;   726 AA.
AC   Q9UW98;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   TRI R 4 ALLERGEN.
OS   Trichophyton rubrum.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Onygenales; Arthrodermataceae; mitosporic Arthrodermataceae;
OC   Trichophyton.
OX   NCBI_TaxID=5551;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99009056; PubMed=9792655;
RA   Woodfolk J.A., Wheatley L.M., Piyasena R.V., Benjamin D.C.,
RA   Platts-Mills T.A.E.;
RT   "Trichophyton antigens associated with IgE antibodies and delayed type
RT   hypersensitivity. Sequence homology to two families of serine
RT   proteinases.";
RL   J. Biol. Chem. 273:29489-29496(1998).
DR   EMBL; AF082514; AAD52012.1; -.
DR   InterPro; IPR000379; -.
SQ   SEQUENCE   726 AA;  80120 MW;  83609D55F06783A4 CRC64;


  Query Match             11.5%;  Score 402.5;  DB 3;  Length 726;
  Best Local Similarity   22.4%;  Pred. No. 1.6e-19;
  Matches  169;  Conservative  120;  Mismatches  245;  Indels  221;  Gaps   34;

Qy     41 SYTSNIWIYETKTGGSVPWTHGE-------KRSTDPRWSPDGRTLAFISDREGDAAQLYI 93
          ::|   :|   : | ::|   ||       | : | :  | |  |  :: : ||   : :
Db     19 AFTPEDFISAPRRGEAIPDPKGELAVFHVSKYNFDKKDRPSGWNL--LNLKNGD---INV 73

Qy     94 MSTEGGEARKLTDIPYGVSKPLWSPDGESILV-----TISLGEGESIDDREKTEQDSYEP 148
          ::|:            ||:  |  ||  ::      ::  | |  | |  |   ::|:
Db     74 LTTDS----------DVSEITWLGDGTKVVYINGTDSVKGGVGIWISD-AKNFGNAYKA 121

Qy    149 VEVQG------------LSYKRDGKGLTRG-----AYAQLVLVSVK------------ 177
            | |             :::   |:  |:|     | |:  : | :
Db    122 GSVNGAFSGLKLAKSGDKINFVGYGQSTTKGDLYNEAAAKEAVSSARIYDSLFVRHWDTY 181

Qy    178 ---------------------SGEMKELTSHK----------ADHGDPAFSPDGKWLVF 205
                              |::| |                ||   | ||| : |
```

```
Db    182 VGTQFNAVFSGTLTKSGDKYSFDGKLKNLVQPVKYAESPYPPFGGSGDYDLSSDGKTVAF 241

Qy    206 --------SANLTETDDASKPHD----VYIMSLESGDLKQVTPH--RGSFGSSSFSPDGR 251
                  |||| :     |||       ::  :|     ||    |:  |  |||||:
Db    242 MSKAPELPKANLTTSYIFLVPHDGSRVAEPINKRNGPR---TPQGIEGASSSPVFSPDGK 298

Qy    252 YLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPI-- 309
           :| |  :   ||        |:         ::   :|:|:            :|:
Db    299 RIAYL--QMAAKN--------YE---------SDRRVIHIAEV---------GTNKPVQR 330

Qy    310 -------------WTKDSQGFYVIGTDQGSTGIYYI-----------------SIEGLVY 339
                      |: | :  ||   |  :  :: :                  |:    :
Db    331 IASNWDRSPEAVKWSSDGRTLYVTAEDHATGKLFTLPADARDNHKPAVVKHDGSVSSFYF 390

Qy    340 PIRLEKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTIS 399
            :   |:  ||  :  : :|:   ||||:           ::|  ||     ||
Db    391 IGSSKSVLISGNSLWSNALYQVAT---PDRPN----------RKLFYAN-----EHDPE 431

Qy    400 I----PEEIQYATEDG--VMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEF- 452
          :    | :|:    ||    :: |:::|   :    |||   |||||   :|  :   :
Db    432 LKGLGPNDIEPLWVDGARTKIHSWIVKPTGFDKNKVYPLAFLIHGGPQGSWGDNWSTRWN 491

Qy    453 -QVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKR 511
           :| | :|| || '|| || |:||:  : :  |:||  | |:::  :        :||
Db    492 PRVWADQGYVVVAPNPTGSTGFGQKLTDDITNDWGGAPYKDLVKIWEHVHDHIKYIDTDN 551

Qy    512 LGVTGGSYGGFMTNWIVGQT--NRFKAAVTQRSISNWISFHGVSDIG---YFFTDWQLEH 566
            | |:|||| ||| ||    :||| |:         :| | | ||    ||    :||
Db    552 GIAAGASFGGFMVNWIQGQDLGRKFKALVSHDG-----TFVGSSKIGTDELFF----IEH 602

Qy    567 D----MFEDTEKL--WDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKE 620
          |     ||  :    || |  :  |   || |::| : | |  :  :   ||  |:: |
Db    603 DFNGTFFEARQNYDRWDCSKPELVAKWSTPQLVVHNDFDFRLSVAEGVGLFNVLQEKGVP 662

Qy    621 TKLVRFPNASHNLSRTGHP---RQRIKRLNYISSW 652
          :: : ||: :| :::  :     |::  | ::: |
Db    663 SRFLNFPDETHWVTKPENSLVWHQQV--LGWVNKW 695


RESULT  15
Q9Y8E3
ID   Q9Y8E3      PRELIMINARY;      PRT;   725 AA.
AC   Q9Y8E3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ALANYL DIPEPTIDYL PEPTIDASE.
OS   Aspergillus oryzae.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5062;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Doumas A., Monod M.;
RT   "Secreted alanyl dipeptidyl peptidase (DppV) of Aspergillus oryzae.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF125190; AAD41777.1; -.
DR   InterPro; IPR000379; -.
SQ   SEQUENCE   725 AA;  80383 MW;  AD5C7FD79F5D666E CRC64;


  Query Match             11.2%;  Score 392.5;  DB 3;  Length 725;
  Best Local Similarity   22.1%;  Pred. No. 7.7e-19;
  Matches  163;  Conservative   95;  Mismatches  255;  Indels  225;  Gaps   28;

Qy     23 PDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRTLAFIS 82
          | |  | : :|| : :    ||   : : |:|                   | : |   |
Db     37 PSGDVAVFSQSQYSFKTHKTTSQWNVLDLKSG------------------DIKLLTNDS 77

Qy     83 D-------REGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESI------------ 123
          |          |:  ||:  |         ||| ||   ||  |
Db     78 DVSEIVWLGSDDSTVLYVNGTN-------ADIPGGVE--LWVSDISDFANGYKAASLPAS 128
```

```
Qy    124 -----LVTISLGE----------------GESIDDREKTEQDSYEPVEVQGLSY------ 156
Db    129 FSGFKVVTTDSGDVRYVAYAESWANGTAYNEELVAKPLSSARIYDSIYVRHWDYYLTTRF 188

Qy    157 ---------KRDGKGLTRGAYAQLVLVSVKSGEMKELTS--HKADHGDPAF--------S 197
Db    189 NAVFSGTLKKSEGKG--KATYK-------ADGDLKNLVSPVKNAESPYPPFGGASDYDLS 239

Qy    198 PDGKWLVFSANLTETDDASKPHDV--------YIMSLESGDLKQVTP------------H 237
Db    240 PDGKWVAFK---------SKAHDIPRANYTTAYIFLVPHDGSKTAVPINGPDSPGTPEGV 290

Qy    238 RGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDI-EQGRLTCLTEMLD-----VHL 291
Db    291 KGDAGSPVFSPDSKKIAYWQMADESYEADHRTLYVYTVGSEETIPSLAADWDRSLDSVKW 350

Qy    292 ADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEY---- 347
Db    351 AD---DDNLIIGVE-----------------DAGRSRLFSI-------PADAGDDYKPKN 383

Qy    348 ------INSFSLSPDEQHFIAS----------VTKPDRPSELYSIPLGQEEKQLTGANDK 391
Db    384 FTDGGVVSAYYQLPDSTYLVTSTAIWTSWNVYIASPEKGVIKTLATANKIDPELKGLGPE 443

Qy    392 FVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHE 451
Db    444 IVDEF---------YYEGNWTKIQAFVIYPENFDKSKSYPLLYYIHGGPQSSWLDSWSTR 494

Qy    452 F--QVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDP 509
Db    495 WNPKVFADQGYVVVAPNPTGSSGFGDALQDAIQNQWGGYPYEDLVKGWEYVNENFDFIDT 554

Qy    510 KRLGVTGGSYGGFMTNWIVGQT--NRFKAAVTQRSISNWISFHGVSDIGYFFTD------ 561
Db    555 DNGVAAGASYGGFMINWIQGSDLGRKFKAL-------------VSHDGTFVADAKVSTE 600

Qy    562 --WQLEHDM---FEDTE---KLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIA 613
Db    601 ELWFMQHEFNGTFWDNRENYRRWDPSAPERILKFSTPMLIIHSDLDYRLPVSEGLSLFNI 660

Qy    614 LKKMGKETKLVRFPNASH 631
Db    661 LQERGVPSRFLNFPDENH 678


Search completed: September 25, 2001, 17:38:52
Job time: 174 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 05:57:53 ; Search time 188.91 Seconds
(without alignments)
6551.237 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 1971
Sequence: 1 atgaaaaagctgataaccgc.....catggtttgatcaacatctc 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1971	100.0	1971	20	AAX07301	Bacillus subtilis
	2	227.2	11.5	1896	19	AAV33582	Pyrococcus horikos
	3	149.8	7.6	1869	19	AAV12887	Thermococcus amida
	4	49.8	2.5	659	21	AAF13721	Aspergillus oryzae
	5	48	2.4	936	22	AAF58252	Oligonucleotide D1
	6	48	2.4	936	22	AAF58254	Oligonucleotide D1
	7	48	2.4	936	22	AAF58257	Oligonucleotide D1
	8	48	2.4	936	22	AAF58259	Oligonucleotide D2
	9	48	2.4	936	22	AAF58262	Oligonucleotide D2
	10	48	2.4	938	22	AAF58255	Oligonucleotide D1
c	11	47.8	2.4	936	22	AAF58252	Oligonucleotide D1
c	12	47.8	2.4	936	22	AAF58254	Oligonucleotide D1
c	13	47.8	2.4	936	22	AAF58257	Oligonucleotide D1
c	14	47.8	2.4	936	22	AAF58259	Oligonucleotide D2
c	15	47.8	2.4	936	22	AAF58262	Oligonucleotide D2
c	16	47.8	2.4	938	22	AAF58255	Oligonucleotide D1
	17	39	2.0	6741	21	AAA10595	Gene encoding a su
	18	37.4	1.9	2196	12	AAQ10958	Gene encoding acyl
c	19	36.2	1.8	9271	19	AAV23080	Fragment HGJ2141 o
	20	35.8	1.8	1395	18	AAT97586	Helicobacter pylor
	21	34.8	1.8	2196	13	AAQ20001	Encodes Human acyl
	22	34.6	1.8	45613	22	AAF28535	Genomic fragment #
c	23	34.2	1.7	729	20	AAZ28861	N.meningitidis nla
	24	34.2	1.7	10732	21	AAA10594	Gene encoding a su
	25	33.6	1.7	420	16	AAQ88233	Bovine aFGF synthe
c	26	33.6	1.7	6151	16	AAQ75977	Insulin receptor s
c	27	33.4	1.7	744	21	AAZ53852	Neisseria meningit
	28	33.4	1.7	957	20	AAZ06344	Cephalosporin acet
	29	33.4	1.7	2277	19	AAV05370	Human telomerase p
	30	33	1.7	4108	21	AAC78099	Human cancer assoc
	31	33	1.7	5301	18	AAT93627	Human metastasis-a
c	32	32.4	1.6	906	21	AAC76786	Human ORFX ORF2341
	33	32.4	1.6	1232	21	AAZ58188	Arabidopsis phosph
	34	32.4	1.6	1242	21	AAC42647	Arabidopsis thalia
	35	32.4	1.6	2277	19	AAV13834	Homo sapiens ambig
	36	32.4	1.6	7680	21	AAX89439	14-3-3 sigma trans
	37	32.2	1.6	2754	20	AAX88420	JP11192089 Seq ID
	38	32.2	1.6	3811	20	AAX38291	E. coli secA DNA.
c	39	32	1.6	438	20	AAV99905	Fragment of homose
c	40	32	1.6	788	20	AAV99902	Fragment of homose
	41	32	1.6	1701	21	AAC49014	Arabidopsis thalia
	42	32	1.6	2366	10	AAN91052	Phage DNA insert o
	43	32	1.6	2381	11	AAQ05241	Sequence enoding a
	44	31.8	1.6	604	22	AAF68340	Human lung tumour
	45	31.8	1.6	2046	12	AAQ14249	Cephalosporin acet

ALIGNMENTS

RESULT 1
AAX07301
ID AAX07301 standard; DNA; 1971 BP.
XX
AC AAX07301;
XX
DT 21-MAY-1999 (first entry)
XX
DE Bacillus subtilis serine protease SP1 (YuxL) DNA.
XX
KW Serine protease SP1; YuxL; detergent; ss.
XX
OS Bacillus subtilis.
XX
PN WO9903984-A2.
XX
PD 28-JAN-1999.
XX
PF 14-JUL-1998; 98WO-US14647.
XX
PR 15-JUL-1997; 97EP-0305232.
XX
PA (GEMV) GENENCOR INT BV.
PA (GEMV) GENENCOR INT INC.
XX
PI Estell DA;
XX
DR WPI; 1999-132231/11.
DR P-PSDB; AAW97789.
XX
PT Use of serine protease genes from Gram-positive microorganisms - for
PT modification of host cells for the production of heterologous
PT proteins or for producing proteins for use in cleaning compositions
XX

PS Disclosure; Fig 1A-C; 37pp; English.
XX
CO This DNA sequence encodes serine protease SP1 (YuxL) of Bacillus
CO subtilis. Novel serine proteases SP1, SP2, SP3, SP4 and SP5 (see
CO AAW97789-93) were identified via a FASTA search of Bacillus subtilis
CO genomic nucleic acid sequences. SP1 was identified by its
CO structural homology to the S9 type serine protease dap2 of yeast.
CO SP2, SP3, SP4 and SP5 were identified by their structural and
CO overall amino acid homology to SP1. Host cells in which the
CO naturally occurring gene encoding one or more of SP1, SP2, SP3,
CO SP4 or SP5 is mutated such that the proteolytic activity is
CO diminished or deleted altogether, can be used for the production of
CO heterologous proteins, e.g. a hormone, enzyme, growth factor,
CO cytokine, protease, carbohydrase, lipase, racemase, epimerase,
CO tautomerase, mutase, transferase, kinase or phosphatase (claimed).
CO SP1, SP2, SP3, SP4 and SP5 can also be produced on a large scale
CO in a microbial host expression system for use in cleaning
CO compositions such as detergents, bar or liquid soap, dish-care
CO formulations and contact lens cleaning solutions, or for peptide
CO hydrolysis, waste treatment, textile applications, as
CO fusion-cleavage enzymes in protein production, and as animal feed
CO additives.
XX
SQ Sequence 1971 BP; 558 A; 431 C; 524 G; 458 T; 0 other;

Query Match 100.0%; Score 1971; DB 20; Length 1971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgaaaaagctgataaccgcagacgacatcacagcgattgtctctgtgaccgatcctcaa 60
Db       1 atgaaaaagctgataaccgcagacgacatcacagcgattgtctctgtgaccgatcctcaa 60

Qy      61 tacgccccagacggtacccgtgccgcatatgtaaaatcacaagtaaatcaagagaaagat 120
Db      61 tacgccccagacggtacccgtgccgcatatgtaaaatcacaagtaaatcaagagaaagat 120

Qy     121 tcgtatacatcaaatatatggatctatgaaacgaaaacgggaggatctgttccttggaca 180
Db     121 tcgtatacatcaaatatatggatctatgaaacgaaaacgggaggatctgttccttggaca 180

Qy     181 catggagaaaaagcgaagcaccgacccaagatggtctccggacgggcgcacgcttgccttt 240
Db     181 catggagaaaagcgaagcaccgacccaagatggtctccggacgggcgcacgcttgccttt 240

Qy     241 atttctgatcgagaaggcgatgcggcacagctttatatcatgagcactgaaggcggagaa 300
Db     241 atttctgatcgagaaggcgatgcggcacagctttatatcatgagcactgaaggcggagaa 300

Qy     301 gcaagaaaactgactgatatcccatatggcgtgtcaaagccgctatggtccccggacggt 360
Db     301 gcaagaaaactgactgatatcccatatggcgtgtcaaagccgctatggtccccggacggt 360

Qy     361 gaatcgattctggtcactatcagtttgggagaggggaaagcattgatgaccgagaaaaa 420
Db     361 gaatcgattctggtcactatcagtttgggagaggggaaagcattgatgaccgagaaaaa 420

Qy     421 acagagcaggacagctatgaacctgttgaagtgcaaggcctctcctacaaacgggacggc 480
Db     421 acagagcaggacagctatgaacctgttgaagtgcaaggcctctcctacaaacgggacggc 480

Qy     481 aaagggctgacgagaggtgcgtatgcccagcttgtgcttgtcagcgtaaagtcgggtgag 540
Db     481 aaagggctgacgagaggtgcgtatgcccagcttgtgcttgtcagcgtaaagtcgggtgag 540

Qy     541 atgaaagagctgacaagtcacaaagctgatcatggtgatcctgctttttctcctgacggc 600
Db     541 atgaaagagctgacaagtcacaaagctgatcatggtgatcctgctttttctcctgacggc 600

Qy     601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660
Db     601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660

Qy     661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720
Db     661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720

Qy     721 ttcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780
Db     721 ttcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780

Qy     781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840
Db     781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840

Qy     841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900
Db     841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900

Qy     901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960
Db     901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960

Qy     961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020
Db     961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020

Qy    1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080
Db    1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080

Qy    1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140
Db    1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140

Qy    1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200
Db    1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200

Qy    1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260
Db    1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260

Qy    1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320
Db    1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320

Qy    1321 catatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggatac 1380
Db    1321 catatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggatac 1380

Qy    1381 gcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatgcg 1440
Db    1381 gcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatgcg 1440

Qy    1441 gtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgaggct 1500
Db    1441 gtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgaggct 1500

Qy    1501 atcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacgga 1560
Db    1501 atcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacgga 1560

Qy    1561 ggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttacccag 1620
Db    1561 ggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttacccag 1620

Qy    1621 cgctcgatatcaaattggatcagctttcacggcgtcagtgatatcggctatttctttaca 1680
Db    1621 cgctcgatatcaaattggatcagctttcacggcgtcagtgatatcggctatttctttaca 1680

Qy    1681 gactggcagcttgagcatgacatgtttgaggacacagaaaagctctgggaccggtctcct 1740
Db    1681 gactggcagcttgagcatgacatgtttgaggacacagaaaagctctgggaccggtctcct 1740
```

```
Qy    1741 ttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatgac 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 ttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatgac 1800

Qy    1801 cgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaaggaa 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 cgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaaggaa 1860

Qy    1861 accaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaaga 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 accaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaaga 1920

Qy    1921 cagcggatcaagcgcctgaattatatcagctcatggtttgatcaacatctc 1971
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 cagcggatcaagcgcctgaattatatcagctcatggtttgatcaacatctc 1971


RESULT   2
AAV33582
ID   AAV33582 standard; DNA; 1896 BP.
XX
AC   AAV33582;
XX
DT   29-DEC-1998  (first entry)
XX
DE   Pyrococcus horikoshi acylpeptide hydrolase encoding DNA sequence.
XX
KW   Pyrococcus horikoshi acylpeptide hydrolase; ss.
XX
OS   Pyrococcus horikoshi.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1896
FT                   /*tag=  a
FT                   /product= "Pyrococcus horikoshi acylpeptide hydrolase"
FT                   /note= "CDS does not contain a stop codon"
XX
PN   JP10210977-A.
XX
PD   11-AUG-1998.
XX
PF   31-JAN-1997;   97JP-0018381.
XX
PR   31-JAN-1997;   97JP-0018381.
XX
PA   (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX
DR   WPI; 1998-488369/42.
DR   P-PSDB; AAW70508.
XX
PT   A heat-resistant acylpeptide hydrolase and a gene coding it - useful
PT   for hydrolysing the C termini of proteins at high temperature
XX
PS   Example 5; Pages 6-7; 8pp; Japanese.
XX
CC   The present sequence represents a Pyrococcus horikoshi acylpeptide
CC   hydrolase encoding DNA sequence.  The enzyme has the following
CC   properties: (a) it hydrolyses an acylpeptide; (b) an optimum temperature
CC   of 90-95 degrees Centigrade; (c) an optimum pH of 5.0-6.0; (d) no loss
CC   of activity when heated to 95 degrees Centigrade for 3 hours at pH 7.5;
CC   and (e) a molecular weight of 60 kDa.  The enzyme is claimed useful
CC   for hydrolysing the amino end of an acylated protein and a peptide at
CC   high temperatures.
XX
SQ   Sequence 1896 BP; 614 A; 358 C; 491 G; 433 T; 0 other;


 Query Match             11.5%;  Score 227.2;  DB 19;  Length 1896;
 Best Local Similarity   55.4%;  Pred. No. 3.3e-62;
 Matches  461;  Conservative    0;  Mismatches  368;  Indels    3;  Gaps    1;

Qy    1139 aggaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaa 1198
           | | | || |  ||  |  |  ||    | |||| |   | | ||   | ||  | ||||

Db    1061 atggaaaggagaagaaggttaccgactttaacaaatggataaagggttacacccttcaa 1120

Qy    1199 tacctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatga 1258
           |||||||  |  || |   |    ||   |||| ||      |  | | |||| | ||||
Db    1121 aacctgaacactttaaggttaaagcaagtgacggggttgaaatagatgcctgggtaatga 1180

Qy    1259 ggcctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtc 1318
           || |   |  |       |  |  |  ||||||  ||||||  | || ||||| ||||
Db    1181 aaccggtgaacttcaggaaaggaaagaagtatccagctattctagagatccacggtggtc 1240

Qy    1319 cgcatatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggat 1378
           |  | |     |||||  |  | | | | || ||||| || ||  |  |  | |||||
Db    1241 ctaaaaccgcttacggttacgcttttatgcacgagttccacgttttaacctctaaaggct 1300

Qy    1379 acgcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatg 1438
           ||  ||  | |     |||||  |||||  |||  | ||||||||  ||||  || | | ||
Db    1301 tcgtcgtgatattctcaaatcctagagggagcgatggctacggagaggagttcgcggat- 1359

Qy    1439 cggtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgagg 1498
           | || ||| | |||||||  | ||||||  | ||| | ||| |||  || |||||
Db    1360 --ataagggacactatgggagagggattaccaggatttaatggaggtagtcgatgaag 1417

Qy    1499 ctatcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacg 1558
           |  | |   ||  |      || |||    |  |||| || || || || ||   ||| |
Db    1418 cattaaggagatttgacttcatagatggggaaaggctaggagttaccgggggttcctatg 1477

Qy    1559 gaggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttaccc 1618
           | || || ||||| |||||||| ||||| || || ||| | || ||||| || || ||||
Db    1478 gtggcttcatgacgaactggatagtcggacataccaacaggttcaaagccgctgtaaccc 1537

Qy    1619 agcgctcgatatcaaattggatcagctttcacggcgtcagtgatatcggctatttcttta 1678
           || | || || |||||||||| |||||   |||    |  ||||| || |||| ||||
Db    1538 agagatcaatttcaaattggataagcttcttcgggacaacggatataggttattactttg 1597

Qy    1679 cagactggcagcttgagcatgacatgtttgaggacacagaaaagctctgggaccggtctc 1738
           |    ||  | |   | ||    |      ||   |||     |||||   |   |
Db    1598 ctccagatcaaataggaaaagatccctggagcaacttggaaggttattgggaaaagagcc 1657

Qy    1739 ctttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatg 1798
           | ||||| |||||  | |||||| || || || ||  | ||| | ||          ||
Db    1658 cattaaagtacgctcccaacgttgaaactcccctgcttataatccactctaccgaagact 1717

Qy    1799 accgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaagg 1858
           || | ||   | |     ||||   ||| || || || ||  | || ||| |  |||| |||
Db    1718 acaggtgttggcttcccgaggcattgcaactcttcatatccctaaaatacctggggaaga 1777

Qy    1859 aaaccaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaa 1918
           |    |  | |    || ||    | |   ||  |  ||    |  | ||  | ||||
Db    1778 gagttgaattggcaatattcccaggagaaaatcatgacctaagtagatctgggaagccaa 1837

Qy    1919 gacagcggatcaagcgcctgaattatatcagctcatggtttgatcaacatct 1970
             ||  || | ||  | ||  |    ||      |||| | ||  ||   ||
Db    1838 agcacagggttaaaagacttgaactaatagcaggatggatggagaaatggct 1889


RESULT   3
AAV12887
ID   AAV12887 standard; DNA; 1869 BP.
XX
AC   AAV12887;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Thermococcus amidase gene.
XX
KW   Amidase; thermostable enzyme; optically active compound;
KW   L-amino acid; peptide; peptidomimetic; archaebacterium; ss.
XX
OS   Thermococcus sp. strain GU5L5.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1869
FT                   /*tag=  a
```

```
FT                   /transl_except= (pos:1228..1230, aa:Val)
FT                   /transl_except= (pos:1231..1233, aa:Val)
XX
PN   WO9748794-A1.
XX
PD   24-DEC-1997.
XX
PF   17-JUN-1997;   97WO-US09319.
XX
PR   17-JUN-1996;   96US-0664646.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Murphy D,  Reid JC,  Robertson D;
XX
DR   WPI; 1998-063135/06.
DR   P-PSDB; AAW41248.
XX
PT   Nucleic acid encoding heat-stable amidase from Thermococcus GU5L5 -
PT   useful for removing N-terminal amino acids from synthetic peptide(s)
PT   and peptidomimetic(s)
XX
PS   Claim 2; Page 39-41; 56pp; English.
XX
CC   This novel isolated nucleic acid encodes a 622 amino acid
CC   thermostable amidase (see AAW41248) of the archaebacterium
CC   Thermococcus GU5L5.  It was isolated from a genomic library of
CC   GU5L5 by screening for amidase activity, and subjecting DNA from a
CC   positive clone to PCR amplification (see AAV12888-89).  The nucleic
CC   acid can be used in the production of amidase in host cells, and to
CC   identify related sequences encoding similar enzymes.  The encoded
CC   amidase is used to remove Arg, Phe or Met from the N-terminus of
CC   synthetic peptides or peptidomimetics (claimed).  Removal of the
CC   N-terminal residue can be done even in the presence of a more
CC   reactive ester bond (very difficult to achieve non-enzymatically).
CC   The amidase is selective for L-amino acids and can therefore be
CC   used to produce optically active compounds.  The protein tolerates
CC   temperatures up to at least 70 degC and high concentrations of
CC   organic solvent (e.g. over 40% dimethyl sulphoxide) so can cleave
CC   bonds that are normally resistant.
XX
SQ   Sequence 1869 BP; 476 A; 486 C; 567 G; 340 T; 0 other;

Query Match             7.6%;  Score 149.8;  DB 19;  Length 1869;
Best Local Similarity   52.3%;  Pred. No. 2.3e-37;
Matches  356;  Conservative    0;  Mismatches  322;  Indels    3;  Gaps    1;

Qy   1291 ccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttcat 1350
           ||  | ||  |   | | ||||||| ||| |   ||||||||||    |  |  |
Db   1135 ccggtgatagtcttcgtccacggcgggccgaagggcatgtacggacaccgcttcgtctac 1194

Qy   1351 gagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagc 1410
          ||| | ||| || |||||   || || |||     |   |   | || ||| | || |||
Db   1195 gagatgcagctgatggcgagcaagggctactactgctgcttcgtgaacccgcgcggcagc 1254

Qy   1411 cacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattat 1470
           |||||||  |  | || || | |      |||   |       |||     |||| | |
Db   1255 gacggctatagcgaagacttcgcgctccgcgtcctggagaggactggcttggaggacttt 1314

Qy   1471 gacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaag 1530
          || ||  | ||| | |   | || |||    |||| |  || |||||    ||     ||
Db   1315 gaggacataatgaacggcatcgaggagttcttcaagctcgaaccgcaggccgacagggag 1374

Qy   1531 cggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcag 1590
          ||  | ||  | ||||||  |||||||||| || || ||||||||||||  | |    |||
Db   1375 cgcgttggaataacgggcataagctacggcggcttcatgaccaactgggccttgactcag 1434

Qy   1591 acgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcac 1650
          |   ||| ||| || || |   | | | ||  |   |||    | ||| ||| |    ||
Db   1435 agcgacctcttcaaggcaggaataagcgagaacggcataagctactggctcaccagctac 1494

Qy   1651 ggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttgag 1710
           | | ||   || || ||    |  |     | |  |   | | ||   |   ||| |||
Db   1495 gccttctcggacatagggctctggtacgacgtcgaggtcatcgggccaaatccgttagag 1554

Qy   1711 gacacagaaaagctctgggaccggtctccttttaaaatacgcagcaaacgtggagacaccg 1770
           ||   || ||  || || | |     ||  |    |||||    |||||| || | |||
Db   1555 aac---gagaacttcaggaagctcagcccgctgttctacgctcagaacgtgaaggcgccg 1611

Qy   1771 cttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctg 1830
           |  |  || |.||         |||  |||| || ||| |||| |||       | |
Db   1612 atactcctaatccactcgcttgaggactaccgctgtccgctcgaccagagccttatgttc 1671

Qy   1831 tttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcg 1890
          |  | ||  || ||  | ||||||||||| || |  | |     ||   |   |   |
Db   1672 tacaacgtgctcaaggacatgggcaaggaagcctacatagcgatattcaagcgcggcgcc 1731

Qy   1891 cacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagc 1950
          |||           |  |||| ||| || || ||  ||||||   |      |||
Db   1732 cacggccacagcgtccgcggaagcccgaggcacaggccgaagcgctacaggctcttcata 1791

Qy   1951 tcatggtttgatcaacatctc 1971
             |  || || |   | |||
Db   1792 gagttcttcgagcgcaagctc 1812


RESULT   4
AAF13721
ID   AAF13721 standard; cDNA; 659 BP.
XX
AC   AAF13721;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Aspergillus oryzae EST SEQ ID NO:6244.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2569; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
```

```
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 659 BP; 162 A; 214 C; 157 G; 126 T; 0 other;


  Query Match             2.5%;  Score 49.8;  DB 21;  Length 659;
  Best Local Similarity   51.1%;  Pred. No. 1.5e-05;
  Matches  117;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy    1358 aggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagccacggct 1417
           ||||  | || |   | |||||||  |||||      ||| || |  ||||||   ||||
Db      47 aggtcttcgccgaccagggatacgtagtcgtcgcacccaacccaaccggaagcagtggct 106

Qy    1418 acgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgatg 1477
           |||  | |   |  | | ||  ||  |    | |  || ||  |    || || ||
Db     107 tcggcgatgccctccaggacgccatccaaaaccaatggggaggctacccctacgaagacc 166

Qy    1478 tgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcg 1537
           |  |  | |    |||  | |  | | |             |||||  |  |       |
Db     167 tcgtcaaaggctgggaatacgtcaacgagaacttcgacttcattgacaccgacaacggtg 226

Qy    1538 gtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgg 1586
             | | | || |  |||||||| || || |||| ||||||||||   ||
Db     227 tcgccgccggtgccagctacggtggcttcatgatcaactggatccaagg 275


RESULT   5
AAF58252
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;


  Query Match             2.4%;  Score 48;  DB 22;  Length 936;
  Best Local Similarity   0.8%;  Pred. No. 7.2e-05;
  Matches    6;  Conservative  410;  Mismatches  340;  Indels    0;  Gaps    0;

Qy     601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660
           ::::   :: :::: : :  :::::::: : ::: : :: ::   :  ::     :: ::
Db      31 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 90

Qy     661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720
            :::: : ::: : : :  : : :  : :: ::::  :  ::: :  : ::      : :
Db      91 wwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 150

Qy     721 ttcggatcaagctcatttttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780
           ::   :: ::  : ::::: :  : :   ::  ::: :: :::  ::  :::: ::::
Db     151 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwww 210

Qy     781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840
           ::::::: :::  ::   : : :::    :   : ::: :::  :: ::       :
Db     211 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 270

Qy     841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900
           : :: : ::: : : ::  :  :| :: :::::    ::    : :::  : ::: :::
Db     271 wwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 330

Qy     901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960
           ::   :  :  : :: :       ::::  : :::  : :   ::   :::::: : ::
Db     331 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 390

Qy     961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020
            ||  | :: ::   : ::    :: :::::::::  ::: ::    :: : :::
Db     391 wgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 450

Qy    1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080
           :::  : :  ::::: : :: :: ::::  :::: : ::: : |: :: :: :  : :::
Db     451 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwww 510

Qy    1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140
           :::   : : : : :::     : : :   : : :  :::: : :::     ::  : :
Db     511 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 570

Qy    1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200
           :: : ::: :  : : :  |  ::: : :: ::: : :   :  :::  :::: ::::
Db     571 wwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 630

Qy    1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260
           : :: : ::: :::::  :: : :: :     : ::  : ::    :   : :: :
Db     631 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwc 690

Qy    1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320
           :  : :::  ::  : : : :: ::::  : ::::: :::: ::: :      :
Db     691 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 750

Qy    1321 catatgatgtacggacatacatattttcatgagttt 1356
            :::: :: ::   : ::: ::::::: :: : :::
Db     751 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 786


RESULT   6
AAF58254
ID   AAF58254 standard; DNA; 936 BP.
XX
```

```
AC   AAF58254;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1875.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

  Query Match             2.4%;  Score 48;  DB 22;  Length 936;
  Best Local Similarity   0.8%;  Pred. No. 7.2e-05;
  Matches    6;  Conservative  410;  Mismatches  340;  Indels    0;  Gaps    0;

Qy     601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660
           ::::   :: :::: : :  :::::::: : ::: : :: ::   :  ::     :: ::
Db      31 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 90

Qy     661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720
           :::: : ::: : : :  : : : : :: ::::  :  ::: :  : ::      : :
Db      91 wwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 150

Qy     721 ttcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780
           ::   :: ::  : ::::: :  : :   ::  ::: ::  :::  ::  :::: ::::
Db     151 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwww 210

Qy     781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840
           ::::::: :::  ::   : : :::    :   : ::: :::  :: ::       :
Db     211 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 270

Qy     841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900
           : :: : ::: : : ::  :  :| :: :::::    ::    : :::  : ::: :::
Db     271 wwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 330

Qy     901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960
           ::   :  :  : :: :        ::::  : :::  : :   ::   :::::: : ::
Db     331 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 390

Qy     961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020
           ||  | :: ::   : ::     :: ::::::::::  ::: ::    :: : :::
Db     391 wgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 450

Qy    1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080
           :::  : :  ::::: : :: :: ::::  :::: : ::: : |: :: :: :  : :::
Db     451 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwww 510

Qy    1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140
           :::   : : : : :::     : : :   : : :  :::: : :::     :: : :
Db     511 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 570

Qy    1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200
           :: : ::: :  : : :   |  ::: : :: ::: : :   :  ::: :::: ::::
Db     571 wwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 630

Qy    1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260
           : :: : ::: :::::  :: : :: :     : ::  : ::    :   : :: :
Db     631 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwc 690

Qy    1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320
           :  : :::: ::  : : : :: :::: : ::::: :::: ::: :  :
Db     691 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 750

Qy    1321 catatgatgtacggacatacatattttcatgagttt 1356
           :::: :: ::   : ::: ::::::: :: : :::
Db     751 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 786


RESULT   7
AAF58257
ID   AAF58257 standard; DNA; 936 BP.
XX
AC   AAF58257;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1954.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

  Query Match             2.4%;  Score 48;  DB 22;  Length 936;
```

```
Best Local Similarity   0.8%;  Pred. No. 7.2e-05;
Matches    6;  Conservative  410;  Mismatches  340;  Indels    0;  Gaps    0;

Qy     601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660
           ::::   :: :::: : :  :::::::: : ::: : :: ::   :  ::     :: ::
Db      31 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 90

Qy     661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720
            :::: : ::: : : :  : : :  : :: :::  :  ::: :  : ::      : :
Db      91 wwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 150

Qy     721 ttcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780
           ::   :: ::  : ::::: :  : :  ::  ::: ::  :::  ::  :::: ::::
Db     151 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwww 210

Qy     781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840
           :::::::: :::  ::   : : :::    :   : ::: :::  :: ::       :
Db     211 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 270

Qy     841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900
           : :: : ::: : : ::  :  :| :: :::::    ::    : :::  : ::: :::
Db     271 wwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 330

Qy     901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960
           ::   :  :  : :: :       ::::  : :::  : :   ::   :::::: : ::
Db     331 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 390

Qy     961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020
            ||  | :: ::   : ::     :: ::::::::::  ::: ::    :: : :::
Db     391 wgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 450

Qy    1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080
           :::  : :  ::::: : :: :: ::::  :::: : ::: : |: :: :: :  : :::
Db     451 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwww 510

Qy    1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140
           :::   : : : :::     : : :   : : :  :::: : :::     ::  : :
Db     511 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 570

Qy    1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200
           :: : ::: :  : : :   |  ::: : :: ::: : :   :  :::  :::: ::::
Db     571 wwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 630

Qy    1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260
            : :: : ::: :::::  :: : :: :     : ::  : ::    :   : :: :
Db     631 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwc 690

Qy    1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320
            :  : ::::  ::  : : : :: ::::  : ::::: :::: ::: :      :
Db     691 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 750

Qy    1321 catatgatgtacggacatacatattttcatgagttt 1356
           :::: :: ::   : ::: ::::::: :: : :::
Db     751 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 786


RESULT   8
AAF58259
ID   AAF58259 standard; DNA; 936 BP.
XX
AC   AAF58259;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2004.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

  Query Match              2.4%;  Score 48;  DB 22;  Length 936;
  Best Local Similarity   0.8%;  Pred. No. 7.2e-05;
  Matches    6;  Conservative  410;  Mismatches  340;  Indels    0;  Gaps    0;

Qy     601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660
           ::::   :: :::: : :  :::::::: : ::: : :: ::   :  ::     :: ::
Db      31 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 90

Qy     661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720
            :::: : ::: : : :  : : :  : :: :::  :  ::: :  : ::      : :
Db      91 wwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 150

Qy     721 ttcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780
           ::   :: ::  : ::::: :  : :  ::  ::: ::  :::  ::  :::: ::::
Db     151 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwww 210

Qy     781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840
           :::::::: :::  ::   : : :::    :   : ::: :::  :: ::       :
Db     211 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 270

Qy     841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900
           : :: : ::: : : ::  :  :| :: :::::    ::    : :::  : ::: :::
Db     271 wwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 330

Qy     901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960
           ::   :  :  : :: :       ::::  : :::  : :   ::   :::::: : ::
Db     331 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 390

Qy     961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020
            ||  | :: ::   : ::     :: ::::::::::  ::: ::    :: : :::
Db     391 wgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 450

Qy    1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080
           :::  : :  ::::: : :: :: ::::  :::: : ::: : |: :: :: :  : :::
Db     451 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwww 510

Qy    1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140
           :::   : : : :::     : : :   : : :  :::: : :::     ::  : :
Db     511 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 570

Qy    1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200
           :: : ::: :  : : :   |  ::: : :: ::: : :   :  :::  :::: ::::
Db     571 wwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 630
```

```
Qy   1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260
            : :: : ::: :::::  :: : :: :     : ::  : ::    :    : :: :
Db    631 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwc 690

Qy   1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320
            :  : ::::  ::  : : : : :: ::::  : ::::: :::: ::: :       :
Db    691 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 750

Qy   1321 catatgatgtacggacatacatattttcatgagttt 1356
           :::: :: ::    : ::: ::::::: :: : :::
Db    751 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 786
```

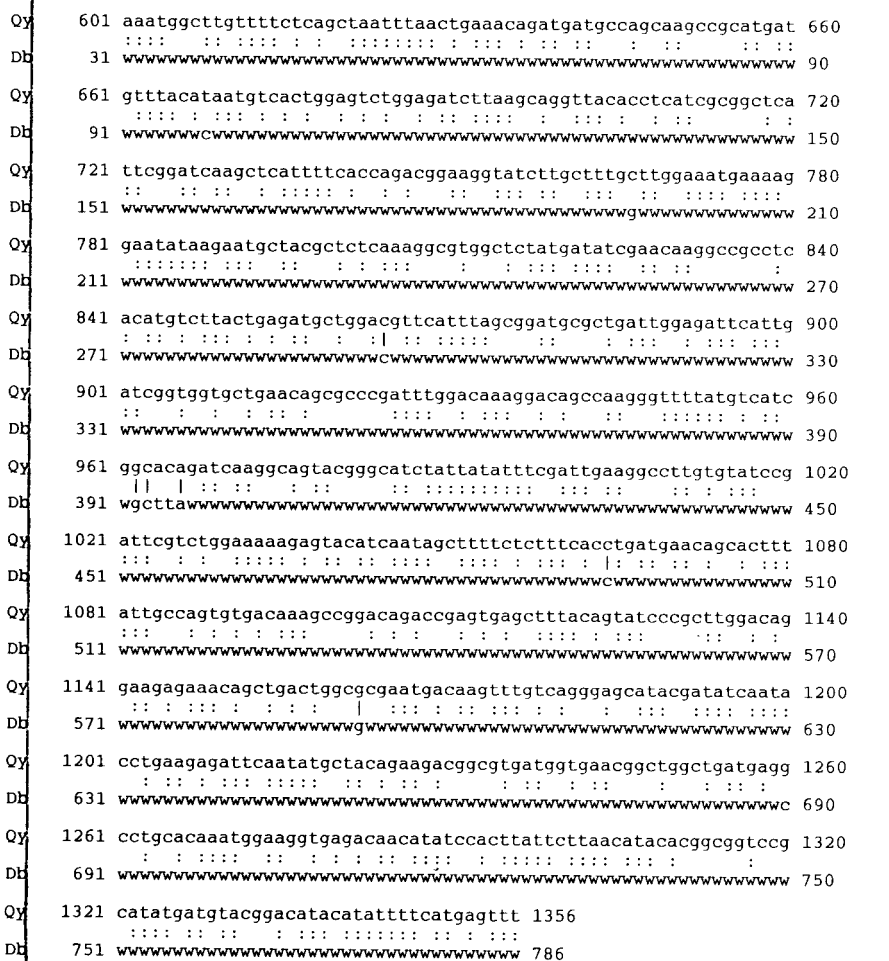
RESULT 9
AAF58262
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry).
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

```
Query Match             2.4%;  Score 48;  DB 22;  Length 936;
Best Local Similarity   0.8%;  Pred. No. 7.2e-05;
Matches    6; Conservative  410; Mismatches  340;  Indels    0;  Gaps    0;

Qy    601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660
           ::::  :: :::: : :  :::::::: : ::: : :: ::    :  ::      :: ::
Db     31 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 90

Qy    661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720
            :::: : ::: : : :  : : :  : :: ::::  :  ::: :  : ::       : :
Db     91 wwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 150

Qy    721 ttcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780
           ::   :: ::  : ::::: :  : :   ::  ::: ::  :::  ::  :::: ::::
Db    151 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwww 210

Qy    781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840
           ::::::: :::  ::   : : :::    :   : ::: ::::  :: ::        :
Db    211 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 270

Qy    841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900
           : :: : ::: : : ::  :  :| :: :::::    ::     : :::  : ::: :::
Db    271 wwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 330

Qy    901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960
           ::   :   :  : :: :        ::::  : :::  : :   ::    :::::: : ::
Db    331 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 390

Qy    961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020
            ||  | :: ::    : ::      :: :::::::::: ::: ::     :: : :::
Db    391 wgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 450

Qy   1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080
           :::   :  :  ::::: : :: :: ::::  :::: : ::: : |: :: :: :  : :::
Db    451 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwww 510

Qy   1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140
           :::    : : : : :::      : : :    : : :  :::: : :::      ::  : :
Db    511 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 570

Qy   1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200
           :: : ::: :   : : :    |  ::: : :: ::: : :    :  :::   :::: ::::
Db    571 wwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 630

Qy   1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260
            : :: : ::: :::::  :: : :: :     : ::  : ::    :    : :: :
Db    631 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwc 690

Qy   1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320
            :  : ::::  ::  : : : : :: ::::  : ::::: :::: ::: :       :
Db    691 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 750

Qy   1321 catatgatgtacggacatacatattttcatgagttt 1356
           :::: :: ::    : ::: ::::::: :: : :::
Db    751 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 786
```

RESULT 10
AAF58255
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX

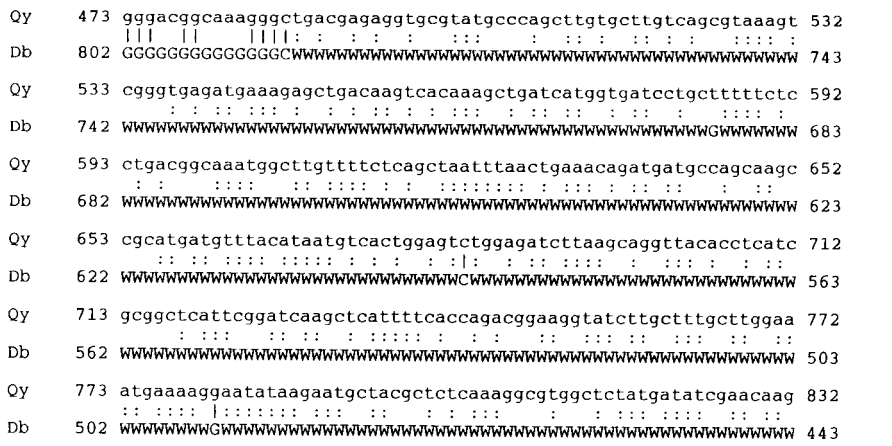
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 2.4%; Score 48; DB 22; Length 938;
Best Local Similarity 0.8%; Pred. No. 7.2e-05;
Matches 6; Conservative 410; Mismatches 340; Indels 0; Gaps 0;

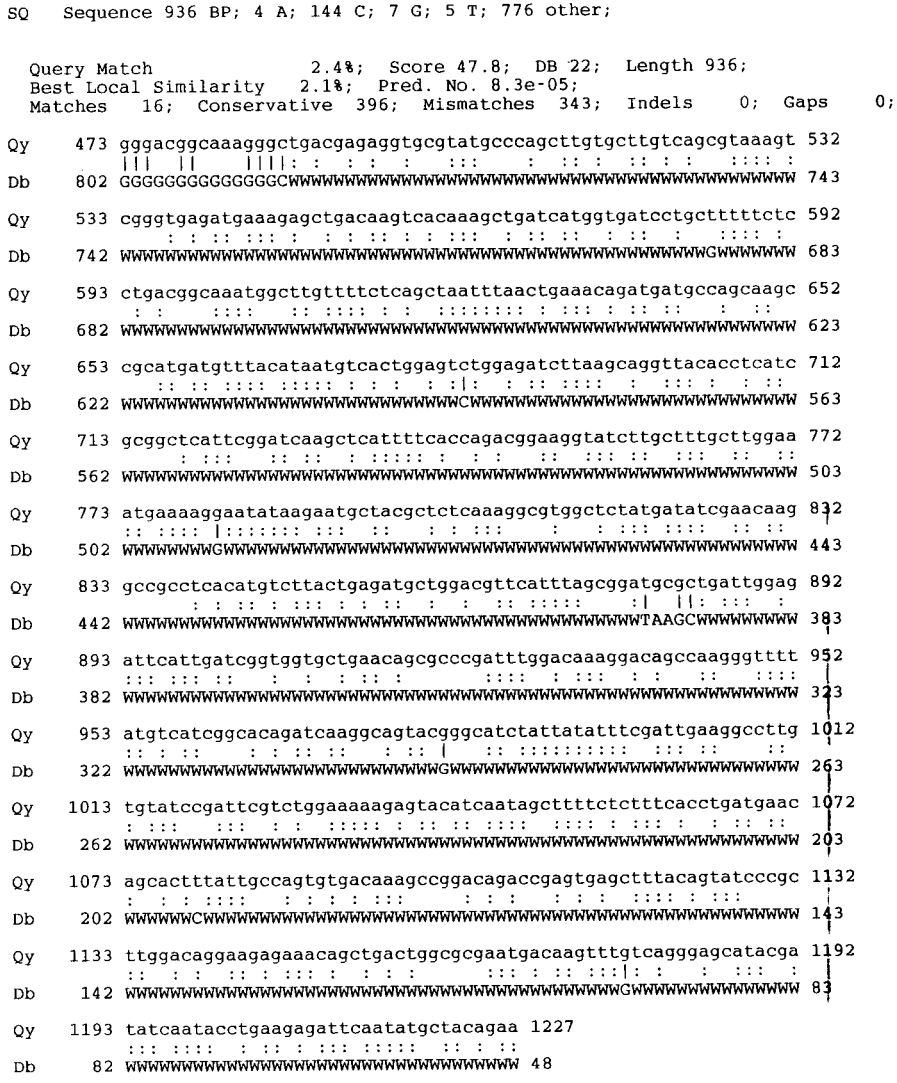


RESULT 11
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 2.4%; Score 47.8; DB 22; Length 936;
Best Local Similarity 2.1%; Pred. No. 8.3e-05;
Matches 16; Conservative 396; Mismatches 343; Indels 0; Gaps 0;



RESULT 12
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX



SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 2.4%; Score 47.8; DB 22; Length 936;
Best Local Similarity 2.1%; Pred. No. 8.3e-05;
Matches 16; Conservative 396; Mismatches 343; Indels 0; Gaps 0;

RESULT 13
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX

OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 2.4%; Score 47.8; DB 22; Length 936;
Best Local Similarity 2.1%; Pred. No. 8.3e-05;
Matches 16; Conservative 396; Mismatches 343; Indels 0; Gaps 0;

Qy 473 gggacggcaaagggctgacgagaggtgcgtatgcccagcttgtgcttgtcagcgtaaagt 532
Db 802 GGGGGGGGGGGGGGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 743

Qy 533 cgggtgagatgaaagagctgacaagtcacaaagctgatcatggtgatcctgctttttctc 592
Db 742 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWW 683

Qy 593 ctgacggcaaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagc 652
Db 682 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 623

Qy 653 cgcatgatgtttacataatgtcactggagtctggagatcttaagcaggttacacctcatc 712
Db 622 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWW 563

Qy 713 gcggctcattcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaa 772
Db 562 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 503

Qy 773 atgaaaaggaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaag 832
Db 502 WWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 443

Qy 833 gccgcctcacatgtcttactgagatgctggacgttcatttagcggatgcgctgattggag 892
Db 442 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWW 383

Qy 893 attcattgatcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggtttt 952
Db 382 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 323

Qy 953 atgtcatcggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttg 1012
Db 322 WWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWW 263

Qy 1013 tgtatccgattcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaaac 1072
Db 262 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 203

Qy 1073 agcactttattgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgc 1132
Db 202 WWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 143

Qy 1133 ttggacaggaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacga 1192
Db 142 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWW 83

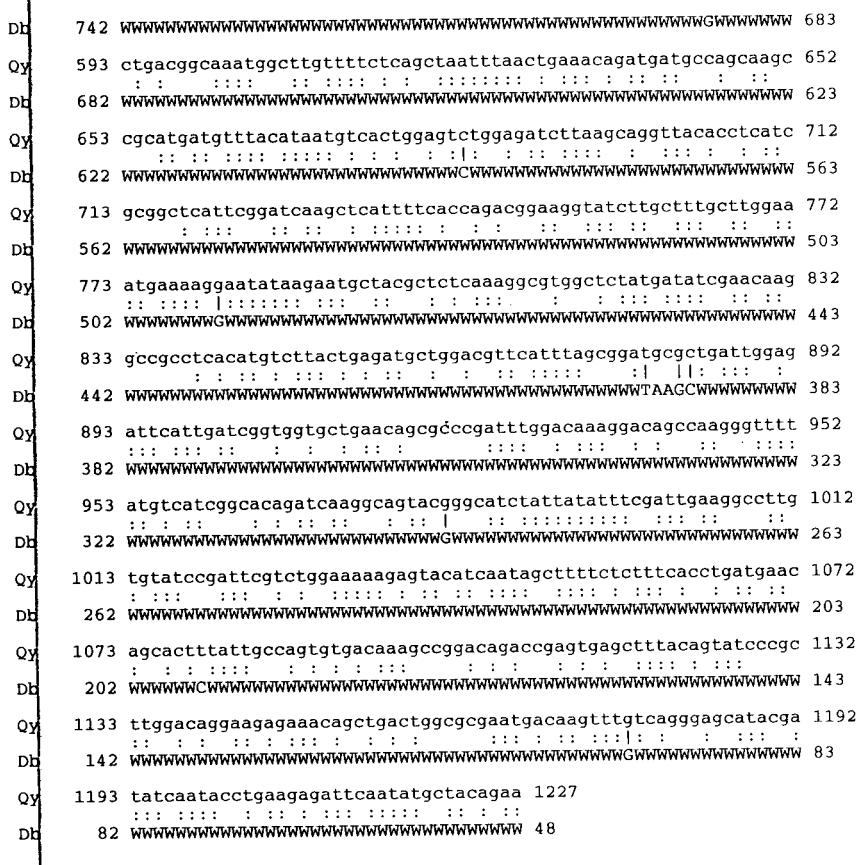
Qy 1193 tatcaatacctgaagagattcaatatgctacagaa 1227
Db 82 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 48

RESULT 14
AAF58259/c
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 2.4%; Score 47.8; DB 22; Length 936;
Best Local Similarity 2.1%; Pred. No. 8.3e-05;
Matches 16; Conservative 396; Mismatches 343; Indels 0; Gaps 0;

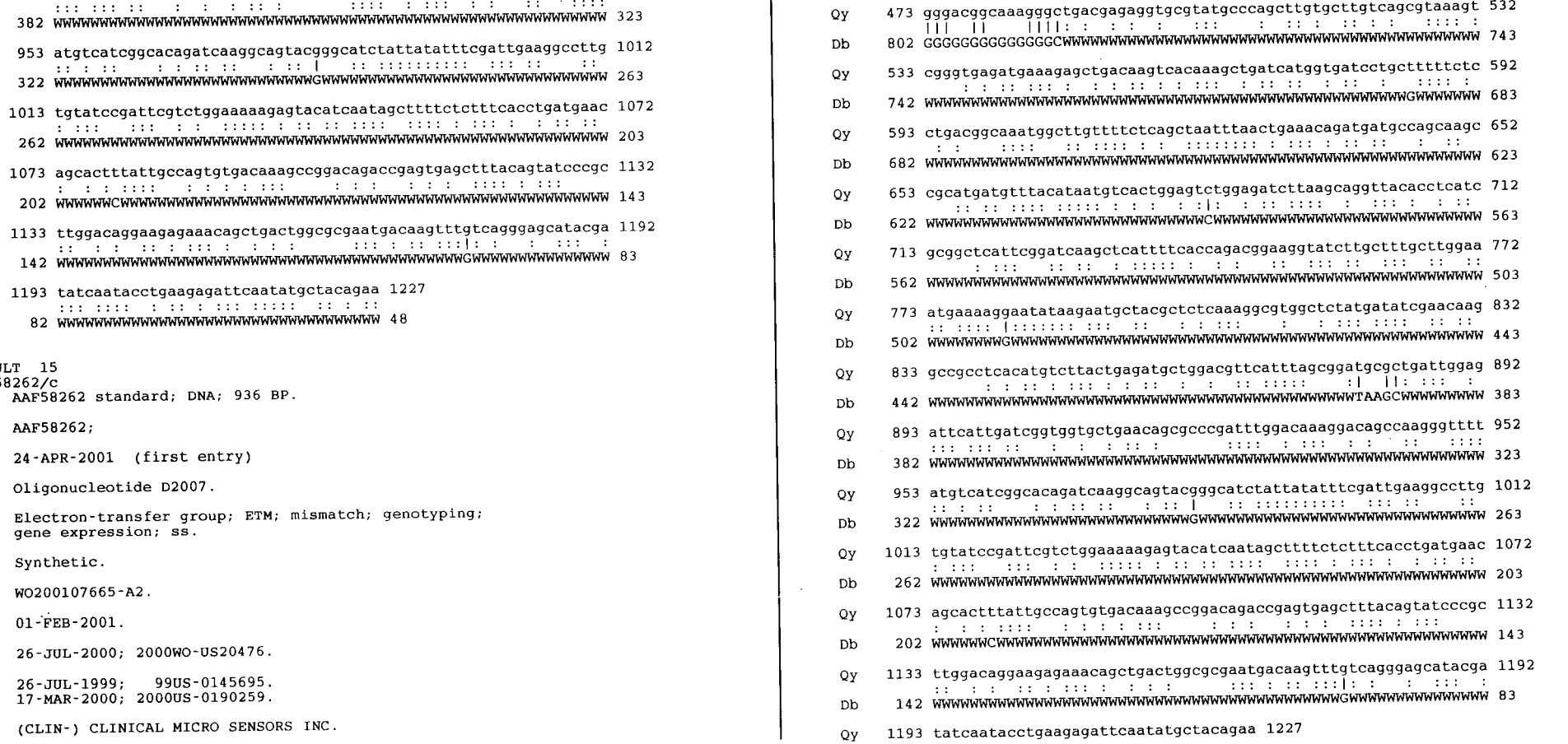
Qy 473 gggacggcaaagggctgacgagaggtgcgtatgcccagcttgtgcttgtcagcgtaaagt 532
Db 802 GGGGGGGGGGGGGGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 743

Qy 533 cgggtgagatgaaagagctgacaagtcacaaagctgatcatggtgatcctgctttttctc 592



RESULT 15
AAF58262/c
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 2.4%; Score 47.8; DB 22; Length 936;
Best Local Similarity 2.1%; Pred. No. 8.3e-05;
Matches 16; Conservative 396; Mismatches 343; Indels 0; Gaps 0;



::: :::: : :: : ::: ::::: :: : ::
Db 82 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 48

Search completed: September 26, 2001, 08:03:30
Job time: 7537 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 02:57:13 ; Search time 2147.34 Seconds
(without alignments)
8676.583 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 1971
Sequence: 1 atgaaaaagctgataaccgc.....catggtttgatcaacatctc 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	66	3.3	839	245	AZ535741	AZ535741 ENTCC88TR
c 2	66	3.3	909	245	AZ530780	AZ530780 ENTBV92TR
c 3	48.4	2.5	327	2	AA071581	AA071581 0028F Pyr
4	46.8	2.4	475	175	BG279712	BG279712 b5e08np.r
c 5	42.6	2.2	873	245	AZ528669	AZ528669 ENTCD42TF
6	42.6	2.2	939	172	BG025910	BG025910 602292057
7	40.8	2.1	1101	219	CNS000D1	AL065414 Drosophil
8	39.4	2.0	711	163	BE130860	BE130860 L48-1811T
9	39.4	2.0	1064	162	BE036593	BE036593 MP01G10 M
c 10	37.4	1.9	1101	219	CNS003EJ	AL064609 Drosophil
11	37.2	1.9	513	122	AW906913	AW906913 EST343036
12	37.2	1.9	884	245	AZ540694	AZ540694 ENTDG10TF
c 13	37.2	1.9	926	248	AZ689494	AZ689494 ENTIN84TF
c 14	37.2	1.9	932	248	AZ692448	AZ692448 ENTHY14TR
15	37.2	1.9	956	219	CNS007BU	AL066969 Drosophil
c 16	37	1.9	459	12	AA792627	AA792627 vs92e09.r
c 17	37	1.9	532	31	AV601962	AV601962 AV601962
c 18	37	1.9	1000	150	BF534027	BF534027 602075007
19	36.8	1.9	377	115	AW353260	AW353260 34766 MAR
20	36.8	1.9	406	115	AW353257	AW353257 34772 MAR
c 21	36.4	1.8	418	31	AV590668	AV590668 AV590668
22	36.4	1.8	457	168	BF707325	BF707325 282867 MA
c 23	36.4	1.8	526	31	AV605765	AV605765 AV605765
24	36.4	1.8	539	31	AV590669	AV590669 AV590669
25	36.4	1.8	549	119	AW657556	AW657556 110968 MA
26	36.4	1.8	550	139	BE755788	BE755788 209719 MA
c 27	36.4	1.8	773	234	AQ856414	AQ856414 nbeb0003G
28	36.2	1.8	621	244	AZ455559	AZ455559 1M0258C01
29	36	1.8	359	115	AW425881	AW425881 58926 MAR
30	36	1.8	377	120	AW786628	AW786628 119969 MA
c 31	36	1.8	929	150	BF580981	BF580981 602100689
c 32	35.8	1.8	987	219	CNS00CBO	AL058976 Drosophil
33	35.6	1.8	627	173	BG097813	BG097813 EST462332
c 34	35.4	1.8	411	32	AV664642	AV664642 AV664642
c 35	35.4	1.8	448	136	BE479402	BE479402 163996 BA
36	35.4	1.8	525	228	AQ400922	AQ400922 HS_5058_A
37	35.2	1.8	615	138	BE660473	BE660473 686 GmaxS
38	35.2	1.8	624	104	AI966343	AI966343 sc37e08.y
39	35.2	1.8	1804	170	BF859397	BF859397 963002C05
40	35	1.8	395	149	BF482398	BF482398 WHE1796_B
41	35	1.8	413	145	BF201425	BF201425 WHE1770_E
42	35	1.8	423	145	BF202692	BF202692 WHE1777_B
43	35	1.8	426	149	BF474305	BF474305 WHE0842_E
c 44	35	1.8	455	171	BF901086	BF901086 IL2-MT017
45	35	1.8	464	136	BE489944	BE489944 WHE0363_F

ALIGNMENTS

RESULT 1
AZ535741/c
LOCUS AZ535741 839 bp DNA GSS 03-NOV-2000
DEFINITION ENTCC88TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ535741
VERSION AZ535741.1 GI:11092688
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 839)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 34
High quality sequence stop: 716.

FEATURES Location/Qualifiers
source 1. .839
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 277 a 143 c 107 g 312 t
ORIGIN

Query Match 3.3%; Score 66; DB 245; Length 839;
Best Local Similarity 48.6%; Pred. No. 8.7e-09;
Matches 214; Conservative 0; Mismatches 220; Indels 6; Gaps 1;

```
Qy   1147 aaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaatacctgaa 1206
          ||||| | |||  |   ||| | |  || |         ||  |   |  | | ||
Db    569 AAACAAATCACTCACATTAATCAAGAAGTTCTTTCTACAATTAAATTTGGAGAATATAAA 510

Qy   1207 gagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgaggcctgca 1266
          || ||||| ||  ||  || | |    |     |  | | |  |  | |    || ||
Db    509 GAAATTCATTACACTGGAGCAAATAATGATCAAATTCATGCATTTATTACTTACCCACCA 450

Qy   1267 caaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccgcatatg 1326
           | ||| ||    | |||| |||||||| | ||| |  | | ||| || || ||  |
Db    449 AATATGAAAACAAACACAAAATATCCAGTAATTTTATATACACATGGAGGACCAGAATCA 390

Qy   1327 atgtacggacatacatattttcat------gagtttcaggtgctggcggcgaaaggatac 1380
               |||  |    | ||       |    || ||  | || ||  |||| ||
Db    389 CCATGGACTAATAATTTCCATTATAGATGGAATCCACAAGTTATAGCAGCACAAGGTTAT 330

Qy   1381 gcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatgcg 1440
            ||  ||     ||||     |||       | || ||  | | |||| | || ||
Db    329 ATTGTTTTTGCACCCAATTTCCATGGATCAGGAAGTTATGGAGATGCATTTTTAAAAGCA 270

Qy   1441 gtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgaggct 1500
           | |||    |||  || |||  |   | ||| ||| | ||| | |   |||||
Db    269 ATTAGAAAGAATTGGGGAGGATGGCCATTTGAAGATTTAATGAAAGGAATGGATTATTTA 210

Qy   1501 atcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacgga 1560
          |  |   | || ||     |||||    |        |||  | ||| | ||| || |||
Db    209 AAAACTAGTGAGCCATTAGTTGATATAGATAATGCATGTGCTATGGGAGCAAGTTATGGA 150

Qy   1561 ggttttatgaccaactggat 1580
           || | |||||  || |||||
Db    149 GGATATATGATGAATTGGAT 130
```

RESULT 2

```
AZ530780/c
LOCUS       AZ530780      909 bp    DNA             GSS       03-NOV-2000
DEFINITION  ENTBV92TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ530780
VERSION     AZ530780.1  GI:11084914
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 909)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 780.
FEATURES             Location/Qualifiers
     source          1. .909
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      299 a    151 c    118 g    341 t
ORIGIN
```

```
  Query Match             3.3%;  Score 66;  DB 245;  Length 909;
  Best Local Similarity   48.6%;  Pred. No. 9e-09;
  Matches  214;  Conservative    0;  Mismatches  220;  Indels    6;  Gaps    1;

Qy    1147 aaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaatacctgaa 1206
           |||||  | |||  |   |||  |  |  || |          ||   |   |  |  | ||
Db     665 AAACAAATCACTCACATTAATCAAGAAGTTCTTTCTACAATTAAATTTGGAGAATATAAA 606

Qy    1207 gagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgaggcctgca 1266
           || ||||| ||  ||  || | |    |       |  | |  |   | |      ||  ||
Db     605 GAAATTCATTACACTGGAGCAAATAATGATCAAATTCATGCATTTATTACTTACCCACCA 546

Qy    1267 caaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccgcatatg 1326
            | ||| ||    | |||| ||||||| | ||| |  | | ||| || || ||  |
Db     545 AATATGAAAACAAACACAAAATATCCAGTAATTTTATATACACATGGAGGACCAGAATCA 486

Qy    1327 atgtacggacatacatattttcat------gagtttcaggtgctggcggcgaaaggatac 1380
             |      |||  |    | ||       |    || ||  | || ||  |||| ||
Db     485 CCATGGACTAATAATTTCCATTATAGATGGAATCCACAAGTTATAGCAGCACAAGGTTAT 426

Qy    1381 gcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatgcg 1440
             ||  ||     ||||      |||       | || ||  | | |||| | || ||
```

```
Db     425 ATTGTTTTTGCACCCAATTTCCATGGATCAGGAAGTTATGGAGATGCATTTTTAAAAGCA 366

Qy    1441 gtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgaggct 1500
            | |||    |||  || |||  |   | ||| ||| | ||| | |   ||||| |
Db     365 ATTAGAAAGAATTGGGGAGGATGGCCATTTGAAGATTTAATGAAAGGAATGGATTATTTA 306

Qy    1501 atcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacgga 1560
            |  |   | || ||     |||||    |        |||  | ||| | ||| || |||
Db     305 AAAACTAGTGAGCCATTAGTTGATATAGATAATGCATGTGCTATGGGAGCAAGTTATGGA 246

Qy    1561 ggtttttatgaccaactggat 1580
            || | |||||  || |||||
Db     245 GGATATATGATGAATTGGAT 226
```

```
RESULT   3
AA071581/c
LOCUS       AA071581      327 bp    mRNA            EST       01-OCT-1996
DEFINITION  0028F Pyrococcus furiosus l-ZAP II library, F Robb Pyrococcus
            furiosus cDNA clone 0028 similar to Acylamine Acid releasing enyzme
            , mRNA sequence.
ACCESSION   AA071581
VERSION     AA071581.1  GI:1578969
KEYWORDS    EST.
SOURCE      Pyrococcus furiosus.
  ORGANISM  Pyrococcus furiosus
            Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
            Pyrococcus.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Borges,K.M., Brummet,S.R., Bogert,A., Davis,M.C., Hujer,K.M., Domke
            ,S.T., Szasz,J., Ravel,J., DiRuggiero,J., Fuller,C., Chase,J.W. and
            Robb,F.T.
  TITLE     A Survey of the Genome of the Hyperthermophilic Archaeon,
            Pyrococcus furiosus
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Brummet SR; Robb F
            Brummet, Sequencing R&D
            Amersham Life Science, Inc.
            Amersham Life Science, 26101 Miles Rd, Cleveland, OH 44128
            Tel: 2164649277; 4102348870
            Fax: 2163600975; 4102348896
            Email: cn288@cleveland.freenet.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Pyrococcus furiosus"
                     /strain="DSM 3638"
                     /db_xref="taxon:2261"
                     /clone="0028"
                     /clone_lib="Pyrococcus furiosus l-ZAP II library, F Robb"
                     /lab_host="E. coli"
                     /note="Vector: l-ZAP II; Site_1: Eco RI; Site_2: Eco RI;
                     Genomic DNA was purified from P. furiosus cells and
                     partially digested with Dra I, Eco RV, Hinc II, Pvu II,
                     size selected, ligated to Eco RI linkers then cloned into
                     the Eco RI site of l-ZAP II, plasmid excision vector.
                     Excision was performed in batch and individual clones
                     retrieved by plating."
BASE COUNT       87 a     82 c     57 g    101 t
ORIGIN
```

```
  Query Match             2.5%;  Score 48.4;  DB 2;  Length 327;
  Best Local Similarity   51.9%;  Pred. No. 0.0016;
  Matches  134;  Conservative    0;  Mismatches  121;  Indels    3;  Gaps    1;

Qy    1239 ggtgaacggctggctgatgaggcctgcacaaatggaaggtgagacaacatatccacttat 1298
           | |  | || |||   |||| ||| |  |  | ||| || ||   || |   ||  | ||
Db     259 GATTGATGGTTGGTATATGAAGCCCGAGCTCAAGGAGGGAGA---AAAAGCCCCCGTGAT 203

Qy    1299 tcttaacatacacggcggtccgcatatgatgtacggacatacatattttcatgagtttca 1358
             |     | || || || ||  |    |||||||| ||   |      ||||  | ||
Db     202 AGTATTTGTCCATGGTGGGCCCAAGGGAATGTACGGTTATTACTTCAAGTATGAAATGCA 143
```

```
Qy    1359 ggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagccacggcta 1418
           |  |||||||  || || |||    | |||||  | |||||  || || ||| | || ||
Db     142 ATTAATGGCGGCCAAGGGGTACTATATAGTTTACGTTAATCCCAGGGGGAGCAATGGATA 83

Qy    1419 cgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgatgt 1478
           |  | || ||||       ||   |        |||     |||| |   | ||| |
Db      82 TAGTGAAGACTTTGCACTAAGAGTATTAACCAGAACTGGACTGGAGGACTTCCAAGATAT 23

Qy    1479 gatgcaggctgtggatga 1496
           |  | | | | || ||
Db      22 ATTAAACGGTATTGAGGA 5


RESULT   4
BG279712
LOCUS       BG279712      475 bp    mRNA            EST       21-FEB-2001
DEFINITION  b5e08np.r1 Neurospora crassa sexual cDNA library, Uni-zap vector
            system Neurospora crassa cDNA clone b5e08np 5', mRNA sequence.
ACCESSION   BG279712
VERSION     BG279712.1  GI:13077352
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Kupfer,D., Lai,H., Nelson,M. and Roe,B.
  TITLE     ESTS from a Neurospora crassa Sexual cDNA Library
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: b5e08np.f1
            Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            Contact Dr. Mary Anne Nelson, Department of Biology, University of
            New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu)
            regarding clone availability
            Seq primer: M13 Universal Reverse Primer
            High quality sequence stop: 474.
FEATURES             Location/Qualifiers
     source          1. .475
                     /organism="Neurospora crassa"
                     /strain="wild type"
                     /db_xref="taxon:5141"
                     /clone="b5e08np"
                     /clone_lib="Neurospora crassa sexual cDNA library, Uni-zap
                     vector system"
                     /tissue_type="perithecia (fruiting bodies)"
                     /dev_stage="sexual"
                     /lab_host="E. coli strain SOLR"
                     /note="Vector: pBlueScript SK-; Site_1: EcoRI; Site_2:
                     XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
                     ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT      103 a    120 c    146 g    105 t      1 others
ORIGIN


  Query Match             2.4%;  Score 46.8;  DB 175;  Length 475;
  Best Local Similarity   52.8%;  Pred. No. 0.0058;
  Matches  122;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy    1360 gtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagccacggctac 1419
           ||  | |||| | |||| ||||  || ||       || |||| ||| ||    || ||
Db     151 GTCTTTGCGGAGCAAGGCTACGTCGTGGTGAGCCCGAACCCGACAGGCAGTACGGGTTAT 210

Qy    1420 gggcaggaatttgtgaa-tgcggtcagaggagattatgggggaaaggattatgacgatgt 1478
           ||   ||   |    ||  |   |||  ||  | |  || || | |   ||||| ||| |
Db     211 GGCATGGCTCTCCAAAACGGNAATCAAGGGCCAGTGGGGTGGCAGGCCGTATGAGGATCT 270

Qy    1479 gatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcgg 1538
           | |  | ||  | ||  |   |    |       ||| |  | ||| |  |  ||   |
Db     271 GGTCAAAGCGTTTGAGCATATTGAGGAGAACATGCCGTACGTAGATACGGACAGGGCGGT 330

Qy    1539 tgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggca 1589
           |    ||| |  || || || |||| ||||| ||| |||||   || ||
Db     331 CGCATTGGGTGCTAGTTATGGTGGTTATATGATCAATTGGATCCAAGGCCA 381


RESULT   5
AZ528669/c
LOCUS       AZ528669      873 bp    DNA             GSS       03-NOV-2000
DEFINITION  ENTCD42TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ528669
VERSION     AZ528669.1  GI:11081093
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 873)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 849.
FEATURES             Location/Qualifiers
     source          1. .873
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      327 a    132 c    101 g    313 t
ORIGIN


  Query Match             2.2%;  Score 42.6;  DB 245;  Length 873;
  Best Local Similarity   54.0%;  Pred. No. 0.14;
  Matches   87;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy    1420 gggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgatgtg 1479
           ||  | | ||||   || ||  | |||    |||  || |||  |   | ||| ||| |
Db     872 GGAGATGCATTTTAAAAAGCAATTAGAAAGAATTGGGGAGGATGGCCATTTGAAGATTTA 813

Qy    1480 atgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcggt 1539
           ||| | |   ||||| |     |      | || ||     |||||    |        ||
Db     812 ATGAAAGGAATGGATTATTTAAAACTTAGTGAGCCATTAGTTGATATAGATAATGCATGT 753
```

```
Qy   1540 gtcacgggcggaagctacggaggttttatgaccaactggat 1580
          |  | ||| | ||| || ||||| | |||||  || |||||
Db    752 GCTATGGGAGCAAGTTATGGAGGATATATGATGAATTGGAT 712


RESULT   6
BG025910
LOCUS       BG025910      939 bp    mRNA            EST       24-JAN-2001
DEFINITION  602292057F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4386711 5',
            mRNA sequence.
ACCESSION   BG025910
VERSION     BG025910.1  GI:12412979
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 939)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10070  row: a  column: 16
            High quality sequence stop: 673.
FEATURES             Location/Qualifiers
     source          1. .939
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4386711"
                     /clone_lib="NIH_MGC_86"
                     /tissue_type="osteosarcoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.533 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      223 a    204 c    251 g    260 t      1 others
ORIGIN


  Query Match             2.2%;  Score 42.6;  DB 172;  Length 939;
  Best Local Similarity   49.1%;  Pred. No. 0.15;
  Matches  111;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy    865 gttcatttagcggatgcgctgattggagattcattgatcggtggtgctgaacagcgcccg 924
          | ||| | |   | |  |  ||  |    |     |||||   ||||||||  |  |  ||
Db    485 GCTCACTGACTCGCTAAGACGACTTTGGCCCTGTTGATGACTGGTGCTGTGCTCCAGCCT 544

Qy    925 atttggacaaaggacagccaagggttttatgtcatcggcacagatcaaggcagtacgggc 984
          |  |  |  ||||          |||  | | | || |||| | || ||   ||  |
Db    545 TATCAGTTAGGGGACCCAAGGTTTGTTTGGGACCTGGGTACAGGTAAAAGCCAGACTTGG 604

Qy    985 atctattatatttcgattgaaggccttgtgtatccgattcgtctggaaaaagagtacatc 1044
          |   |   | |||| |       |||| ||  ||  | | | ||| | |  | |  |||
Db    605 AGGAACCCTCTTTCTAGGCTGAACCTTGAGTCCCCCCTGCTTTTGGCAGACTAATGGATC 664

Qy   1045 aatagcttttctctttcacctgatgaacagcactttattgccagtg 1090
          | |  |||    ||    | | |||    |  || | | |||||||
Db    665 ACTGTCTTGCAGCTAGTNCTTCATGTTGGGGCCTCTTTAGCCAGTG 710


RESULT   7
CNS000D1
LOCUS       CNS000D1     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065414
VERSION     AL065414.1  GI:4938827
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR01J16"
                     /note="end : TET3"
BASE COUNT      280 a    104 c    123 g    211 t    383 others
ORIGIN


  Query Match             2.1%;  Score 40.8;  DB 219;  Length 1101;
  Best Local Similarity   14.9%;  Pred. No. 0.56;
  Matches   58;  Conservative  171;  Mismatches  161;  Indels    0;  Gaps    0;

Qy   1135 ggacaggaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgata 1194
           :|| |:| | |:|:|  : ::: : : :    |      |||:  | :::: ::
Db    692 KGAGARGKATAKARAKKRWKDDKDTKKKKTKTTATKKKTTTTKRAATDKWKWRKKKATAG 751

Qy   1195 tcaatacctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctg 1254
             ::::   |: |:  :: ::|  : :      :   | : :  ||:|:: :|  || ::
Db    752 AWDWDDARRGRGGRWGDKRRRTGARGDKDKTGKKKKTGDKRDKGGGDGDRGDGGAGGGDR 811

Qy   1255 atgaggcctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggc 1314
           : :|:   :  :  :  |  :: :: :| :   ::|    : :|  ::|: |:: : ::
Db    812 DRDARARRARWDTKDTRGAKDKADKKKAGRKRKWWTRKTWKRDTADWKAWAAWRWRWRRD 871

Qy   1315 ggtccgcatatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaa 1374
           :::  | :::::::: | ::: ::: :||::::  :: ::|  |     ::: :: :: :
Db    872 RKDRRGRRDRWRRWRGADKKWKWDDDDTAKDDWTKKDTDWTAGATAGWDDDKAKDGDWKR 931

Qy   1375 ggatacgcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtg 1434
           | :::   : : |:    :: : |  :||:|:: :    :: |:  :  :|:: : : |:
Db    932 GTDWRTWKRTWKGWGRKGDKGRKTAGRAGDGKDKKKGKDKDRTDDAKATDGKDTWRDTTD 991

Qy   1435 aatgcggtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggat 1494
            |:: ::: ::: |:|::   :::::::|: ::  ::  ::::::: : :|: : : :|
Db    992 TAKRRRRKRRRRAGRGRDRKGDKKKRWRARWRKGKKKKTKDKDDDDGKRWGRAKAKADAA 1051

Qy   1495 gaggctatcaaacgagatccgcatattgat 1524
           :|::  :  :||  |:|:  : ::|:::|:
```

```
Db    1052 KAKDDADAADAAAWAKAKAKDKDKAKKKAD 1081


RESULT   8
BE130860
LOCUS       BE130860      711 bp    mRNA            EST       20-FEB-2001
DEFINITION  L48-1811T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1811
            5', mRNA sequence.
ACCESSION   BE130860
VERSION     BE130860.1  GI:8578249
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
            PCR PRimers
            FORWARD: T7
            BACKWARD: T3
            Plate: L48-19  row: A  column: 11
            Seq primer: T3
            High quality sequence stop: 350
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone="L48-1811"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression library
                     , 48 hours NaCl treatment"
                     /tissue_type="Leaf, 48 h 0.4M NaCl"
                     /dev_stage="Six week old"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT      170 a    159 c    158 g    224 t
ORIGIN


  Query Match             2.0%;  Score 39.4;  DB 163;  Length 711;
  Best Local Similarity   54.5%;  Pred. No. 1.3;
  Matches   79;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     867 tcatttagcggatgcgctgattggagattcattgatcggtggtgctgaacagcgcccgat 926
           | || |   |||| | | ||||| | |||| ||||| |  |     |||  ||| || |
Db     382 TAATGTTATGGATACCCAGATTGCATATTCCTTGATTGAAGAACAGGAAGGGCGACCAAG 441

Qy     927 ttggacaaaggacagccaagggttttatgtcatcggcacagatcaaggcagtacgggcat 986
           | | || | |||         |||  ||   ||    | |||| | ||  |   |||||
Db     442 ATTGCCAGATGACTATATCTCCTTTGTTGCTCTCCTTGCTGATCCACGCTATTGTGGCAT 501

Qy     987 ctattatatttcgattgaaggcctt 1011
           | |||| ||  ||  || |   ||
Db     502 TTCTTATCTTGAGAAGGAGGAGGTT 526


RESULT   9
BE036593
LOCUS       BE036593     1064 bp    mRNA            EST       07-JUN-2000
DEFINITION  MP01G10 MP Mesembryanthemum crystallinum cDNA 5', mRNA sequence.
ACCESSION   BE036593
VERSION     BE036593.1  GI:8331602
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 1064)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            Best blastx match: 'gi|3643594 (AC005395) unknown protein
            [Arabidopsis thaliana] 316 2e-85'. An open reading frame exists.
            Insert Length: 1   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .1064
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone_lib="MP"
                     /tissue_type="apical meristem and leaf primordia"
                     /dev_stage="6 weeks"
                     /note="3 d 500mM NaCl"
BASE COUNT      286 a    220 c    230 g    311 t     17 others
ORIGIN


  Query Match             2.0%;  Score 39.4;  DB 162;  Length 1064;
  Best Local Similarity   54.5%;  Pred. No. 1.5;
  Matches   79;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     867 tcatttagcggatgcgctgattggagattcattgatcggtggtgctgaacagcgcccgat 926
           | || |   |||| | | ||||| | |||| ||||| |  |     |||  ||| || |
Db     312 TAATGTTATGGATACCCAGATTGCATATTCCTTGATTGAAGAACAGGAAGGGCGACCAAG 371

Qy     927 ttggacaaaggacagccaagggttttatgtcatcggcacagatcaaggcagtacgggcat 986
           | | || | |||         |||  ||   ||    | |||| | ||  |   |||||
Db     372 ATTGCCAGATGACTATATCTCCTTTGTTGCTCTCCTTGCTGATCCACGCTATTGTGGCAT 431

Qy     987 ctattatatttcgattgaaggcctt 1011
           | |||| ||  ||  || |   ||
Db     432 TTCTTATCTTGAGAAGGAGGAGGTT 456


RESULT  10
CNS003EJ/c
LOCUS       CNS003EJ     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08I02 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064609
VERSION     AL064609.1  GI:4941961
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
```

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08I02"
                     /note="end : TET3"
BASE COUNT      270 a    257 c    203 g    324 t     47 others
ORIGIN
```

```
  Query Match              1.9%;  Score 37.4;  DB 219;  Length 1101;
  Best Local Similarity   59.6%;  Pred. No. 6.2;
  Matches   59;  Conservative    2;  Mismatches   38;  Indels    0;  Gaps    0;

Qy     388 ggagaggggaaagcattgatgaccgagaaaaaacagagcaggacagctatgaacctgtt 447
           || |: |: ||||| |   | | |||| ||  ||  ||||| ||||   | ||| | | |
Db     905 GGGGRAGRAGAAAGGAGGAAGGGCCGAAAAGGAAGGGAGCAAGACACAAACGAAGCAGAT 846

Qy     448 gaagtgcaaggcctctcctacaaacgggacggcaaaggg 486
           |||| | || | ||    ||  ||  | |  ||| ||||
Db     845 GAAGAGGAATGGCTAAATTAAGAAATGCAGTGCAGAGGG 807
```

```
RESULT  11
AW906913
LOCUS       AW906913      513 bp    mRNA            EST       24-MAY-2000
DEFINITION  EST343036 potato stolon, Cornell University Solanum tuberosum cDNA
            clone cSTA4N22, mRNA sequence.
ACCESSION   AW906913
VERSION     AW906913.1  GI:8071123
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
            Holt,I.E., Liang,F., Hansen,T.S., Utterback,T., Bowman,C.L., Doan
            ,B., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker
            ,B.
  TITLE     Generation of ESTs from potato swelling stolons
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Research Genetics, Libraries Division
            Tel: 1-800-711-6195
            Email: cdna@resgen.com
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Solanum tuberosum"
                     /cultivar="Bintje"
                     /db_xref="taxon:4113"
                     /clone="cSTA4N22"
                     /clone_lib="potato stolon, Cornell University"
                     /tissue_type="axillary buds of stem explants, swelling
                     stolons"
                     /dev_stage="1 to 3 days"
                     /lab_host="SOLR"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; RNA was supplied by Christian Bachem & Beatrix
                     Horvath(Laboratory of Plant Breeding, Dept. of Plant
                     Sciences,  Wageningen University, The Netherlands). Total
                     RNA was isolated from developing axillary buds of potato
                     nodal stem cuttings  cultured on medium for the
                     introduction of tuber formation as  described in Bachem et
                     al. (Plant Journal 1996). Tissue samples were  taken of
                     stages corresponding to growing stolons and the early
                     stages  of tuber formation."
BASE COUNT      117 a    106 c    122 g    168 t
ORIGIN
```

```
  Query Match              1.9%;  Score 37.2;  DB 122;  Length 513;
  Best Local Similarity   48.2%;  Pred. No. 5.4;
  Matches  105;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    1289 atccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttc 1348
           |||| || ||| |   | | ||||| ||||| |||    | |        |  |  |
Db     242 ATCCGCTGATTGTAGTCCTTCACGGGGGTCCTCATTCTGTTTCATTGTCAAGCTTCTCAA 301

Qy    1349 atgagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaa 1408
           |   ||     | ||  |  |    || ||    |  |   | | |||   |||||
Db     302 AGTCCTTGGCTTTCCTTTCTTCACTTGGTTATAGCTTGTTGATTGTAAATTATAGAGGCT 361

Qy    1409 gccacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggatt 1468
           |   || |  ||| |||||   ||  |  | |    ||  |   |||   | |||||
Db     362 CCTTGGGTTTTGGGGAGGAAGCAGTACAATCTCTTCCTGGGAAAATTGGATCACAGGATG 421

Qy    1469 atgacgatgtgatgcaggctgtggatgaggctatcaaa 1506
           | | ||||||  |    ||| | ||| | |  ||  ||
Db     422 TTAATGATGTGCTTGCTGCTATAGATCATGTCATTGAA 459
```

```
RESULT  12
AZ540694
LOCUS       AZ540694      884 bp    DNA             GSS       14-NOV-2000
DEFINITION  ENTDG10TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ540694
VERSION     AZ540694.1  GI:11147729
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 27
            High quality sequence stop: 844.
FEATURES             Location/Qualifiers
     source          1. .884
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
```

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

BASE COUNT 315 a 89 c 226 g 254 t
ORIGIN

Query Match 1.9%; Score 37.2; DB 245; Length 884;
Best Local Similarity 50.9%; Pred. No. 6.6;
Matches 115; Conservative 0; Mismatches 108; Indels 3; Gaps 1;

```
Qy    1364 tggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagccacggctacgggc 1423
           ||||  |  |||| |||    |  ||      |||  ||||||  |  || |  ||
Db     101 TGGCATCAGAAGGGTACATTATTATTGCACCGAATAGAAGAGGATTACCAGGTTTTGGAA 160

Qy    1424 aggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgatgtgatgc 1483
           |||||   |  | |   |   |  |||||||| ||      |    | ||| || |
Db     161 TGGAATGGTTAGAAGAAATTTCATTAGATTATGGAGGTCTTTGTATGAAGGATTTGTTAA 220

Qy    1484 aggctgtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcggtgtca 1543
             ||||   |||    || |||  |||  |  |  ||||| || |  | || ||
Db     221 CTTCTGT---TGACATAATGAAAGAAGAGTCATACGTTGATTCTGACAGACTTGGATGTG 277

Qy    1544 cgggcggaagctacggaggttttatgaccaactggatcgtcgggca 1589
             || | ||| |  ||||| | |  |    | ||| | |  |||||
Db     278 TTGGAGCAAGTTTTGGAGGATATTCGGTTTATTGGCTTGCTGGGCA 323
```

RESULT 13
AZ689494/c
LOCUS AZ689494 926 bp DNA GSS 14-DEC-2000
DEFINITION ENTIN84TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ689494
VERSION AZ689494.1 GI:11826640
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 926)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 18
High quality sequence stop: 862.
FEATURES Location/Qualifiers
source 1. .926
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

BASE COUNT 266 a 218 c 110 g 332 t
ORIGIN

Query Match 1.9%; Score 37.2; DB 248; Length 926;
Best Local Similarity 50.9%; Pred. No. 6.7;
Matches 115; Conservative 0; Mismatches 108; Indels 3; Gaps 1;

```
Qy    1364 tggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagccacggctacgggc 1423
           ||||  |  |||| |||    |  ||      |||  ||||||  |  || |  ||
Db     714 TGGCATCAGAAGGGTACATTATTATTGCACCGAATAGAAGAGGATTACCAGGTTTTGGAA 655

Qy    1424 aggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgatgtgatgc 1483
           |||||   |  | |   |   |  |||||||| ||      |    | ||| || |
Db     654 TGGAATGGTTAGAAGAAATTTCATTAGATTATGGAGGTCTTTGTATGAAGGATTTGTTAA 595

Qy    1484 aggctgtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcggtgtca 1543
             ||||   |||    || |||  |||  |  |  ||||| || |  | || ||
Db     594 CTTCTGT---TGACATAATGAAAGAAGAGTCATACGTTGATTCTGACAGACTTGGATGTG 538

Qy    1544 cgggcggaagctacggaggttttatgaccaactggatcgtcgggca 1589
             || | ||| |  ||||| | |  |    | ||| | |  |||||
Db     537 TTGGAGCAAGTTTTGGAGGATATTCGGTTTATTGGCTTGCTGGGCA 492
```

RESULT 14
AZ692448/c
LOCUS AZ692448 932 bp DNA GSS 14-DEC-2000
DEFINITION ENTHY14TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ692448
VERSION AZ692448.1 GI:11829594
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 932)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 454.
FEATURES Location/Qualifiers
source 1. .932
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

```
BASE COUNT      209 a    201 c     70 g    452 t
ORIGIN
```

```
  Query Match             1.9%;  Score 37.2;  DB 248;  Length 932;
  Best Local Similarity   59.4%;  Pred. No. 6.7;
  Matches   63;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     685 ggagatcttaagcaggttacacctcatcgcggctcattcggatcaagctcattttcacca 744
           ||| || ||| | |   ||    |  ||||||   |||     || || ||||| |||||
Db     888 GGAAATTTTAGGGAATATAAGGGTTTTCGCGGGAAATTGAACCCAGGCCCATTTGCACCA 829

Qy     745 gacggaaggtatcttgctttgcttggaaatgaaaaggaatataaga 790
           || | ||| ||| |  | ||   | | || |   |  |||||||
Db     828 GAAGCAAGTTATATGCCATTTACTTGGAAAGCCTAATAATATAATA 783
```

```
RESULT  15
CNS007BU
LOCUS       CNS007BU      956 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR15O23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066969
VERSION     AL066969.1  GI:4945533
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 956)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR15O23"
                     /note="end : TET3"
BASE COUNT      259 a    146 c    201 g    223 t    127 others
ORIGIN
```

```
  Query Match             1.9%;  Score 37.2;  DB 219;  Length 956;
  Best Local Similarity   31.5%;  Pred. No. 6.8;
  Matches   56;  Conservative   48;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     886 attggagattcattgatcggtggtgctgaacagcgcccgatttggacaaaggacagccaa 945
           | |:|| ||: :| |:|   ||| :  :||   ::  :|  ||    ||::   || :||
Db     726 ASTSGAAATKGRTAGRTWATTGGARGGRAAGGASBTTSGGSTTTTAAAVKTTTAGASAA 785

Qy     946 gggttttatgtcatcggcacagatcaaggcagtacgggcatctattatatttcgattgaa 1005
           ::  | : | : :|::::| | |:            :| || : ||:|:|  :: | |
Db     786 SSAATGKTTTWAWTSSKSAAAAAWTKSTTTTWRTKTKGAATTBTTTVTWTSGSRGTAGGG 845

Qy    1006 ggccttgtgtatccgattcgtctggaaaaagagtacatcaatagcttttctctttcac 1063
           ::: |:  | : ::: |: ::    :|||| : | |: || |  :||: | ||:::
Db     846 SSMATDMSGCVVSSVTTWAKKGRAWWAAAAGGRAAAAWWAAAAWTWTTKTTTTTKSVS 903
```

Search completed: September 26, 2001, 06:56:28
Job time: 14355 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 03:09:38 ; Search time 86.18 Seconds
(without alignments)
4329.685 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 1971
Sequence: 1 atgaaaaagctgataaccgc.....catggtttgatcaacatctc 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	227.2	11.5	1896	3	US-09-016-080-2	Sequence 2, Appli
	2	149.8	7.6	1869	2	US-08-664-646A-1	Sequence 1, Appli
	3	149.8	7.6	1869	2	US-09-066-285-1	Sequence 1, Appli
	4	149.8	7.6	1869	3	US-09-261-006-1	Sequence 1, Appli
	5	149.8	7.6	1869	3	US-08-951-088-1	Sequence 1, Appli
c	6	54.4	2.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	7	35.8	1.8	1395	3	US-08-834-776A-1	Sequence 1, Appli
c	8	34.2	1.7	729	2	US-08-512-955-3	Sequence 3, Appli
c	9	33.6	1.7	6152	1	US-08-557-139-1	Sequence 1, Appli
	10	33.4	1.7	2277	1	US-08-676-967-2	Sequence 2, Appli
	11	33.4	1.7	2277	1	US-08-676-974-2	Sequence 2, Appli
	12	33.4	1.7	2277	2	US-09-098-487-2	Sequence 2, Appli
	13	31.8	1.6	957	1	US-07-688-299-2	Sequence 2, Appli
	14	31.8	1.6	957	1	US-07-980-517A-2	Sequence 2, Appli
	15	31.8	1.6	2046	1	US-07-688-299-12	Sequence 12, Appl
	16	31.8	1.6	2046	1	US-07-980-517A-11	Sequence 11, Appl
c	17	31.4	1.6	2743	1	US-08-396-479B-3	Sequence 3, Appli
c	18	31.4	1.6	2743	1	US-08-818-823-3	Sequence 3, Appli
c	19	31.4	1.6	2749	2	US-08-124-981A-1	Sequence 1, Appli
c	20	31.4	1.6	2751	3	US-09-037-190-45	Sequence 45, Appl
c	21	31.4	1.6	2751	3	US-09-037-192-45	Sequence 45, Appl
c	22	31.4	1.6	2751	3	US-09-037-143-45	Sequence 45, Appl
c	23	31.4	1.6	2751	4	US-09-049-691-45	Sequence 45, Appl
c	24	31.4	1.6	2751	4	US-08-260-174-45	Sequence 45, Appl
c	25	31.4	1.6	2853	5	PCT-US94-07297-36	Sequence 36, Appl
c	26	31.2	1.6	1348	2	US-08-743-637B-15	Sequence 15, Appl
c	27	31.2	1.6	1348	3	US-08-526-840B-15	Sequence 15, Appl
	28	31.2	1.6	4928	1	US-08-399-561-1	Sequence 1, Appli
c	29	31	1.6	1193	1	US-07-956-697B-4	Sequence 4, Appli
c	30	31	1.6	1193	1	US-08-263-098-4	Sequence 4, Appli
	31	31	1.6	4248	3	US-08-678-614-1	Sequence 1, Appli
c	32	30.6	1.6	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	33	30.6	1.6	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	34	30.6	1.6	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	35	30	1.5	1044	3	US-08-513-974B-47	Sequence 47, Appl
c	36	30	1.5	1607	3	US-08-513-974B-341	Sequence 341, App
	37	29.8	1.5	1168	4	US-09-029-213B-5	Sequence 5, Appli
	38	29.8	1.5	1423	4	US-08-916-576B-3	Sequence 3, Appli
	39	29.8	1.5	5820	4	US-09-029-213B-7	Sequence 7, Appli
	40	29.8	1.5	35081	2	US-08-752-760A-1	Sequence 1, Appli
	41	29.6	1.5	1366	3	US-09-191-099-4	Sequence 4, Appli
	42	29.6	1.5	1413	3	US-09-191-099-1	Sequence 1, Appli
	43	29.6	1.5	1462	3	US-09-191-099-3	Sequence 3, Appli
	44	29.4	1.5	1176	3	US-09-072-384-16	Sequence 16, Appl
c	45	29.4	1.5	2018	3	US-08-714-918-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-016-080-2
; Sequence 2, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kaxuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1896
; TYPE: DNA
; ORGANISM: Pyrococcus horikoshii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1896)
US-09-016-080-2

Query Match 11.5%; Score 227.2; DB 3; Length 1896;
Best Local Similarity 55.4%; Pred. No. 1.2e-67;
Matches 461; Conservative 0; Mismatches 368; Indels 3; Gaps 1;

```
Qy   1139 aggaagagaaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaa 1198
          | | | || |  ||  |  | ||    | |||| |   | | ||   | ||  | ||||
Db   1061 atggaaaggagaagaaggttaccgactttaacaaatggataaagggttacacccctttcaa 1120

Qy   1199 tacctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatga 1258
          ||||||| |  || |   |    ||   ||||| ||      |  | | |||| | ||||
Db   1121 aacctgaacactttaaggttaaagcaagtgacggggttgaaatagatgcctgggtaatga 1180

Qy   1259 ggcctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtc 1318
             | |   | |       |  | |  | ||||||  ||||||  | || |||||  ||||
Db   1181 aaccggtgaacttcaggaaaggaaagaagtatccagctattctagagatccacggtggtc 1240

Qy   1319 cgcatatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggat 1378
          |  | |     |||||  |  | | | | || ||||| || ||  |  |  | ||||| |
Db   1241 ctaaaaccgcttacggttacgcttttatgcacgagttccacgttttaacctctaaaggct 1300

Qy   1379 acgcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatg 1438
```

```
Db   1301 tcgtcgtgatattctcaaatcctagagggagcgatggctacggagaggagttcgcggat- 1359

Qy   1439 cggtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgagg 1498

Db   1360 --ataaggggacactatggggagagggattaccaggatttaatggaggtagtcgatgaag 1417

Qy   1499 ctatcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacg 1558

Db   1418 cattaaggagatttgacttcatagatgggggaaaggctaggagttaccgggggttcctatg 1477

Qy   1559 gaggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttaccc 1618

Db   1478 gtggcttcatgacgaactggatagtcggacataccaacaggttcaaagccgctgtaaccc 1537

Qy   1619 agcgctcgatatcaaattggatcagctttcacggcgtcagtgatatcggctatttcttta 1678

Db   1538 agagatcaatttcaaattggataagcttcttcgggacaacggatataggttattactttg 1597

Qy   1679 cagactggcagcttgagcatgacatgtttgaggacacagaaaagctctgggaccggtctc 1738

Db   1598 ctccagatcaaataggaaaagatccctggagcaacttggaaggttattgggaaaagagcc 1657

Qy   1739 ctttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatg 1798

Db   1658 cattaaagtacgctcccaacgttgaaactcccctgcttataatccactctaccgaagact 1717

Qy   1799 accgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaagg 1858

Db   1718 acaggtgttggcttcccgaggcattgcaactcttcatatccctaaaatacctggggaaga 1777

Qy   1859 aaaccaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaa 1918

Db   1778 gagttgaattggcaatattcccaggagaaaatcatgacctaagtagatctgggaagccaa 1837

Qy   1919 gacagcggatcaagcgcctgaattatatcagctcatggtttgatcaacatct 1970

Db   1838 agcacagggttaaaagacttgaactaatagcaggatggatggagaaatggct 1889
```

```
RESULT   2
US-08-664-646A-1
; Sequence 1, Application US/08664646A
; Patent No. 5877001
; GENERAL INFORMATION:
;   APPLICANT:  Murphy et al.
;   TITLE OF INVENTION:  Amidases
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;     ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;     STREET:  6 BECKER FARM ROAD
;     CITY:  ROSELAND
;     STATE:  NEW JERSEY
;     COUNTRY:  USA
;     ZIP:  07068
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 INCH DISKETTE
;     COMPUTER:  IBM PS/2
;     OPERATING SYSTEM:  MS-DOS
;     SOFTWARE:  WORD PERFECT 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/664,646A
;     FILING DATE:  June 17, 1996
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Charles J. Herron
;     REGISTRATION NUMBER:  28,019
;     REFERENCE/DOCKET NUMBER:  331400-53
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  201-994-1700
;     TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1869 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-08-664-646A-1
```

```
  Query Match              7.6%;  Score 149.8;  DB 2;  Length 1869;
  Best Local Similarity   52.3%;  Pred. No. 5.6e-41;
  Matches  356;  Conservative    0;  Mismatches  322;  Indels    3;  Gaps    1;

Qy   1291 ccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttcat 1350

Db   1135 CCGGTGATAGTCTTCGTCCACGGCGGGCCGAAGGGCATGTACGGACACCGCTTCGTCTAC 1194

Qy   1351 gagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagc 1410

Db   1195 GAGATGCAGCTGATGGCGAGCAAGGGCTACTACTGCTGCTTCGTGAACCCGCGCGGCAGC 1254

Qy   1411 cacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattat 1470

Db   1255 GACGGCTATAGCGAAGACTTCGCGCTCCGCGTCCTGGAGAGGACTGGCTTGGAGGACTTT 1314

Qy   1471 gacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaag 1530

Db   1315 GAGGACATAATGAACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCAGGCCGACAGGGAG 1374

Qy   1531 cggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcag 1590

Db   1375 CGCGTTGGAATAACGGGCATAAGCTACGGCGGCTTCATGACCAACTGGGCCTTGACTCAG 1434

Qy   1591 acgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcac 1650

Db   1435 AGCGACCTCTTCAAGGCAGGAATAAGCGAGAACGGCATAAGCTACTGGCTCACCAGCTAC 1494

Qy   1651 ggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttgag 1710

Db   1495 GCCTTCTCGGACATAGGGCTCTGGTACGACGTCGAGGTCATCGGGCCAAATCCGTTAGAG 1554

Qy   1711 gacacagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccg 1770

Db   1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTGTTCTACGCTCAGAACGTGAAGGCGCCG 1611

Qy   1771 cttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctg 1830

Db   1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCCGCTCGACCAGAGCCTTATGTTC 1671

Qy   1831 tttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcg 1890

Db   1672 TACAACGTGCTCAAGGACATGGGCAAGGAAGCCTACATAGCGATATTCAAGCGCGGCGCC 1731

Qy   1891 cacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagc 1950

Db   1732 CACGGCCACAGCGTCCGCGGAAGCCCGAGGCACAGGCCGAAGCGCTACAGGCTCTTCATA 1791

Qy   1951 tcatggtttgatcaacatctc 1971

Db   1792 GAGTTCTTCGAGCGCAAGCTC 1812
```

```
RESULT   3
US-09-066-285-1
; Sequence 1, Application US/09066285
; Patent No. 5985646
; GENERAL INFORMATION:
;   APPLICANT:  Murphy et al.
;   TITLE OF INVENTION:  Amidases
;   NUMBER OF SEQUENCES:  4
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/066,285
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/664,646
;      FILING DATE:  June 17, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-53
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1869 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-09-066-285-1


  Query Match             7.6%;  Score 149.8;  DB 2;  Length 1869;
  Best Local Similarity   52.3%;  Pred. No. 5.6e-41;
  Matches  356;  Conservative    0;  Mismatches  322;  Indels    3;  Gaps    1;

Qy   1291 ccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttcat 1350
          ||  | ||  |   | | |||||||| ||| |    ||||||||||    |   | |
Db   1135 CCGGTGATAGTCTTCGTCCACGGCGGGCCGAAGGGCATGTACGGACACCGCTTCGTCTAC 1194

Qy   1351 gagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagc 1410
          ||| | ||| || |||||   || || |||     |   |   | || ||| | || |||
Db   1195 GAGATGCAGCTGATGGCGAGCAAGGGCTACTACTGCTGCTTCGTGAACCCGCGCGGCAGC 1254

Qy   1411 cacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattat 1470
          |||||||  |  | || || | |      |||   |      |||     |||| | |
Db   1255 GACGGCTATAGCGAAGACTTCGCGCTCCGCGTCCTGGAGAGGACTGGCTTGGAGGACTTT 1314

Qy   1471 gacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaag 1530
          || ||  | ||| | |   | || |||    |||| |  || |||||    ||     ||
Db   1315 GAGGACATAATGAACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCAGGCCGACAGGGAG 1374

Qy   1531 cggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcag 1590
          ||  | ||  | ||||||  |||||||| || || ||||||||||||  | |    |||
Db   1375 CGCGTTGGAATAACGGGCATAAGCTACGGCGGCTTCATGACCAACTGGGCCTTGACTCAG 1434

Qy   1591 acgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcac 1650
          |   ||| ||| || || |   | | | ||  |  |||    | ||| ||| |    ||
Db   1435 AGCGACCTCTTCAAGGCAGGAATAAGCGAGAACGGCATAAGCTACTGGCTCACCAGCTAC 1494

Qy   1651 ggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttgag 1710
          | | ||   || || ||    |  |      | |  |  | ||   |   ||| |||
Db   1495 GCCTTCTCGGACATAGGGCTCTGGTACGACGTCGAGGTCATCGGGCCAAATCCGTTAGAG 1554

Qy   1711 gacacagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccg 1770
           ||   || ||  || || | |     ||  |    |||||    |||||| || | |||
Db   1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTGTTCTACGCTCAGAACGTGAAGGCGCCG 1611

Qy   1771 cttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctg 1830
          |  |  || | ||          |||  |||| || ||| |||| |||        | |
Db   1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCCGCTCGACCAGAGCCTTATGTTC 1671

Qy   1831 tttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcg 1890
          |  | ||  || ||  | ||||||||||||| || |  | |     ||   |   |  |
Db   1672 TACAACGTGCTCAAGGACATGGGCAAGGAAGCCTACATAGCGATATTCAAGCGCGGCGCC 1731

Qy   1891 cacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagc 1950
          |||           |  ||||  ||| || ||  ||   ||||||   |      |||
Db   1732 CACGGCCACAGCGTCCGCGGAAGCCCGAGGCACAGGCCGAAGCGCTACAGGCTCTTCATA 1791

Qy   1951 tcatggtttgatcaacatctc 1971
             |  || ||  |   | |||
Db   1792 GAGTTCTTCGAGCGCAAGCTC 1812


RESULT   4
US-09-261-006-1
; Sequence 1, Application US/09261006
; Patent No. 6004796
;  GENERAL INFORMATION:
;    APPLICANT:  Murphy et al.
;    TITLE OF INVENTION:  Amidases
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/261,006
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/664,646
;      FILING DATE:  June 17, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-53
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1869 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-09-261-006-1


  Query Match             7.6%;  Score 149.8;  DB 3;  Length 1869;
  Best Local Similarity   52.3%;  Pred. No. 5.6e-41;
  Matches  356;  Conservative    0;  Mismatches  322;  Indels    3;  Gaps    1;

Qy   1291 ccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttcat 1350
          ||  | ||  |   | | |||||||| ||| |    ||||||||||    |   | |
Db   1135 CCGGTGATAGTCTTCGTCCACGGCGGGCCGAAGGGCATGTACGGACACCGCTTCGTCTAC 1194

Qy   1351 gagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagc 1410
          ||| | ||| || |||||   || || |||     |   |   | || ||| | || |||
Db   1195 GAGATGCAGCTGATGGCGAGCAAGGGCTACTACTGCTGCTTCGTGAACCCGCGCGGCAGC 1254
```

```
Qy   1411 cacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattat 1470
Db   1255 GACGGCTATAGCGAAGACTTCGCGCTCCGCGTCCTGGAGAGGACTGGCTTGGAGGACTTT 1314

Qy   1471 gacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaag 1530
Db   1315 GAGGACATAATGAACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCAGGCCGACAGGGAG 1374

Qy   1531 cggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcag 1590
Db   1375 CGCGTTGGAATAACGGGCATAAGCTACGGCGGCTTCATGACCAACTGGGCCTTGACTCAG 1434

Qy   1591 acgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcac 1650
Db   1435 AGCGACCTCTTCAAGGCAGGAATAAGCGAGAACGGCATAAGCTACTGGCTCACCAGCTAC 1494

Qy   1651 ggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttgag 1710
Db   1495 GCCTTCTCGGACATAGGGCTCTGGTACGACGTCGAGGTCATCGGGCCAAATCCGTTAGAG 1554

Qy   1711 gacacagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccg 1770
Db   1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTGTTCTACGCTCAGAACGTGAAGGCGCCG 1611

Qy   1771 cttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctg 1830
Db   1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCCGCTCGACCAGAGCCTTATGTTC 1671

Qy   1831 tttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcg 1890
Db   1672 TACAACGTGCTCAAGGACATGGGCAAGGAAGCCTACATAGCGATATTCAAGCGCGGCGCC 1731

Qy   1891 cacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagc 1950
Db   1732 CACGGCCACAGCGTCCGCGGAAGCCCGAGGCACAGGCCGAAGCGCTACAGGCTCTTCATA 1791

Qy   1951 tcatggtttgatcaacatctc 1971
Db   1792 GAGTTCTTCGAGCGCAAGCTC 1812
```

```
RESULT   5
US-08-951-088-1
; Sequence 1, Application US/08951088
; Patent No. 6136583
; GENERAL INFORMATION:
;   APPLICANT:  Murphy et al.
;   TITLE OF INVENTION:  Amidases
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;     ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;     STREET:  6 BECKER FARM ROAD
;     CITY:  ROSELAND
;     STATE:  NEW JERSEY
;     COUNTRY:  USA
;     ZIP:  07068
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 INCH DISKETTE
;     COMPUTER:  IBM PS/2
;     OPERATING SYSTEM:  MS-DOS
;     SOFTWARE:  WORD PERFECT 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/951,088
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/664,646
;     FILING DATE:  June 17, 1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Charles J. Herron
;     REGISTRATION NUMBER:  28,019
;     REFERENCE/DOCKET NUMBER:  331400-53
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  201-994-1700
;     TELEFAX:  201-994-1744
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1869 NUCLEOTIDES
;     TYPE:  NUCLEIC ACID
;     STRANDEDNESS:  SINGLE
;     TOPOLOGY:  LINEAR
;   MOLECULE TYPE:  DNA
US-08-951-088-1

  Query Match             7.6%;  Score 149.8;  DB 3;  Length 1869;
  Best Local Similarity   52.3%;  Pred. No. 5.6e-41;
  Matches  356;  Conservative    0;  Mismatches  322;  Indels    3;  Gaps    1;

Qy   1291 ccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttcat 1350
Db   1135 CCGGTGATAGTCTTCGTCCACGGCGGGCCGAAGGGCATGTACGGACACCGCTTCGTCTAC 1194

Qy   1351 gagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagc 1410
Db   1195 GAGATGCAGCTGATGGCGAGCAAGGGCTACTACTGCTGCTTCGTGAACCCGCGCGGCAGC 1254

Qy   1411 cacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattat 1470
Db   1255 GACGGCTATAGCGAAGACTTCGCGCTCCGCGTCCTGGAGAGGACTGGCTTGGAGGACTTT 1314

Qy   1471 gacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaag 1530
Db   1315 GAGGACATAATGAACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCAGGCCGACAGGGAG 1374

Qy   1531 cggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcag 1590
Db   1375 CGCGTTGGAATAACGGGCATAAGCTACGGCGGCTTCATGACCAACTGGGCCTTGACTCAG 1434

Qy   1591 acgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcac 1650
Db   1435 AGCGACCTCTTCAAGGCAGGAATAAGCGAGAACGGCATAAGCTACTGGCTCACCAGCTAC 1494

Qy   1651 ggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttgag 1710
Db   1495 GCCTTCTCGGACATAGGGCTCTGGTACGACGTCGAGGTCATCGGGCCAAATCCGTTAGAG 1554

Qy   1711 gacacagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccg 1770
Db   1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTGTTCTACGCTCAGAACGTGAAGGCGCCG 1611

Qy   1771 cttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctg 1830
Db   1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCCGCTCGACCAGAGCCTTATGTTC 1671

Qy   1831 tttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcg 1890
Db   1672 TACAACGTGCTCAAGGACATGGGCAAGGAAGCCTACATAGCGATATTCAAGCGCGGCGCC 1731

Qy   1891 cacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagc 1950
Db   1732 CACGGCCACAGCGTCCGCGGAAGCCCGAGGCACAGGCCGAAGCGCTACAGGCTCTTCATA 1791

Qy   1951 tcatggtttgatcaacatctc 1971
Db   1792 GAGTTCTTCGAGCGCAAGCTC 1812
```

```
RESULT   6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT:  DORNER, F.
```

```
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             2.8%;  Score 54.4;  DB 1;  Length 7218;
  Best Local Similarity   4.8%;  Pred. No. 1.1e-07;
  Matches   19;  Conservative  219;  Mismatches  160;  Indels    0;  Gaps    0;

Qy      100 caagtaaatcaagagaaagattcgtatacatcaaatatatggatctatgaaacgaaaacg 159
            ||||||  | ||||||  |||    | |    |::: : : :::    : :::: ::::: :
Db     1466 CAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1407

Qy      160 ggaggatctgttccttggacacatggagaaaagcgaagcaccgacccaagatggtctccg 219
            ::::::   :      ::: : : ::::::::: :::: :  ::   :::: ::     :
Db     1406 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1347

Qy      220 gacgggcgcacgcttgcctttatttctgatcgagaaggcgatgcggcacagctttatatc 279
            :: ::: : : :   :     :     ::  ::::::: :: : :: : ::    : :
Db     1346 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1287

Qy      280 atgagcactgaaggcggagaagcaagaaaactgactgatatcccatatggcgtgtcaaag 339
            : ::: :  ::::: ::::::: :::::::  ::  :: :    : : :: : :  :::
Db     1286 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1227

Qy      340 ccgctatggtccccggacggtgaatcgattctggtcactatcagtttgggagaggggaa 399
            :  : ::     ::: :: :::  ::    ::  :  :  ::   :::::::::::::
Db     1226 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1167

Qy      400 agcattgatgaccgagaaaaaacagagcaggacagctatgaacctgttgaagtgcaaggc 459
            :: :  :: ::  ::::::::: :::: :::: ::  : :::   :  :::: : ::::
Db     1166 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1107

Qy      460 ctctcctacaaacgggacggcaaagggctgacgagagg 497
```

```
                : ::: :::: :: ::::::  :: ::::::
Db     1106 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069


RESULT   7
US-08-834-776A-1
; Sequence 1, Application US/08834776A
; Patent No. 6060241
;  GENERAL INFORMATION:
;    APPLICANT:  Corthesy-Theulaz, Irene
;    TITLE OF INVENTION:  Compositions and Methods Relating to
;    TITLE OF INVENTION:  Drug Discovery and Detection and Treatment of
;    TITLE OF INVENTION:  Gastrointestinal Diseases
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
;      STREET:  Five Palo Alto Square, 3000 El Camino Real
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306-2155
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/834,776A
;      FILING DATE:  03-APR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ashton, Nina M.
;      REGISTRATION NUMBER:  37,273
;      REFERENCE/DOCKET NUMBER:  GAST-001/01US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5000
;      TELEFAX:  415-857-0663
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;      LENGTH:  1395 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-834-776A-1


  Query Match             1.8%;  Score 35.8;  DB 3;  Length 1395;
  Best Local Similarity   46.9%;  Pred. No. 0.085;
  Matches  112;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy      898 ttgatcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtc 957
            ||  | || |||  ||  | | ||  |||       | | | ||   |     |||||
Db      136 TTAGTGGGCGGTTTTGGGCTGTGCGGGATACCCGAATACGCCATTGATTACATTTATAAG 195

Qy      958 atcggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtat 1017
            |  ||||   |  |    | |   |  | | || ||  |    |||| |   ||| |  |
Db      196 AAAGGCATTAAGGATTTGATTGTCGTGAGCAATAATTGTGGCGTTGATGATTTTGGGCTT 255

Qy     1018 ccgattcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcac 1077
              |||| | | |||||| || | |||||  |   ||     |||   ||  | | |  |
Db      256 GGCATTCTTTTAGAAAAAAAGCAGATCAAAAAGATTATCGCTTCGTATGTGGGAGAAAAT 315

Qy     1078 tttattgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttgg 1136
              |||        |  || || | || ||   | |||  |       |  ||||  ||   |
Db      316 AAGATTTTTGAATCGCAAATGCTGAACGGAGAAATTGAAGTCGTTTTGACACCGCAAGG 374


RESULT   8
US-08-512-955-3/c
; Sequence 3, Application US/08512955
; Patent No. 5976536
;  GENERAL INFORMATION:
```

```
;   APPLICANT:  Stephens, David S.
;   APPLICANT:  Kahler, Charlene M.
;   TITLE OF INVENTION:  Neisseria Mutants, Lipooligosaccharides
;   TITLE OF INVENTION:  and Immunogenic Compositions
;   NUMBER OF SEQUENCES:  5
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Greenlee and Winner, P.C.
;     STREET:  5370 Manhattan Circle, Suite 201
;     CITY:  Boulder
;     STATE:  Colorado
;     COUNTRY:  US
;     ZIP:  80303
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/512,955
;     FILING DATE:  09-AUG-1995
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Ferber, Donna M.
;     REGISTRATION NUMBER:  33,878
;     REFERENCE/DOCKET NUMBER:  12-95
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (303) 499-8080
;     TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  729 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   ORIGINAL SOURCE:
;     ORGANISM:  Neisseria meningitidis
;     STRAIN:  NMB
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  1..729
US-08-512-955-3


  Query Match             1.7%;  Score 34.2;  DB 2;  Length 729;
  Best Local Similarity   58.3%;  Pred. No. 0.19;
  Matches   60;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy    1418 acgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgatg 1477
           ||||   ||   | | | |||   || | |  ||| ||| || |   ||    |||||||
Db     206 ACGGCGGGGCGGTCGGGGATGTTTTCCGCGCCGATGATGCGGTATTTGAGCCCGACGATG 147

Qy    1478 tgatgcaggctgtggatgaggctatcaaacgagatccgcatat 1520
           || |  ||   | || ||||| | |  | | ||| ||||  ||
Db     146 TGTTTGAGCGAGAGGTTGAGGATTTTGACCCAGACCCGCGCAT 104


RESULT  9
US-08-557-139-1/c
; Sequence 1, Application US/08557139
; Patent No. 5827730
;  GENERAL INFORMATION:
;    APPLICANT:  Pedersen, Oluf
;    APPLICANT:  Bjorbak, Christian
;    APPLICANT:  Frederiksen, Kathrine A.
;    TITLE OF INVENTION:  MUTANT DNA ENCODING INSULIN RECEPTOR
;    TITLE OF INVENTION:  SUBSTRATE 1
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5827730o No. 5827730disk of No. 5827730th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/557,139
;      FILING DATE:  12-FEB-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4041.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 867-0123
;      TELEFAX:  (212) 878-9655
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6152 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  581..4309
US-08-557-139-1


  Query Match             1.7%;  Score 33.6;  DB 1;  Length 6152;
  Best Local Similarity   59.4%;  Pred. No. 1.5;
  Matches   57;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     426 gcaggacagctatgaacctgttgaagtgcaaggcctctcctacaaacgggacggcaaagg 485
           | |||| |||   | ||| | ||   |||  |||||| |   |   || || |||||||
Db     914 GTAGGAGAGCCTGGTACCAGCTGTCTTGCTCGGCCTCGCTGTCCGCCGCGATGGCAAAGT 855

Qy     486 gctgacgagaggtgcgtatgcccagcttgtgcttgt 521
           |||       ||||   |   ||| |  ||||||||
Db     854 GCTCGTCCCGGGTGTAGAGAGCCACCAGGTGCTTGT 819


RESULT  10
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,967
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
```

```
;       REFERENCE/DOCKET NUMBER:  UCB96-055
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415)343-4341
;       TELEFAX:  (415)343-4342
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2277 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-676-967-2


  Query Match             1.7%;  Score 33.4;  DB 1;  Length 2277;
  Best Local Similarity   27.3%;  Pred. No. 0.82;
  Matches   81;  Conservative   52;  Mismatches  164;  Indels    0;  Gaps    0;

Qy     386 tgggagaggggaaagcattgatgaccgagaaaaaacagagcaggacagctatgaacctg 445
           ||| :||:| :||:| :  :|||:||:   ||  |    ||  |:||   :   |    :|
Db     683 TGGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGAYGAYGARGARGAYGGNGTNTTYG 742

Qy     446 ttgaagtgcaaggcctctcctacaaacgggacggcaaagggctgacgagaggtgcgtatg 505
           :|| | : | |         | ||    ||  : ||:|     | :   |   :  :
Db     743 AYGAYGARGAYGARGARGARGARAAYATHGARWSNAARGTNACNAARCCNGTNCARATHC 802

Qy     506 cccagcttgtgcttgtcagcgtaaagtcgggtgagatgaaagagctgacaagtcacaaag 565
             |::  |   |    :|        | |   |::  :  || |  ::     |  |
Db     803 ARAARMGNGCNGTNAARMGNCCNGCNCCNGCNAARWSNWSNGAYCAYWSNGARGARGAYW 862

Qy     566 ctgatcatggtgatcctgctttttctcctgacggcaaatggcttgttttctcagctaatt 625
           : ||::  |  ||  : | ::    :  :||:||  |:  ::| |     :: | :|  :
Db     863 SNGAYYTNGARGARWSNGAYWSNATHGAYGAYGGNGARGARYTNGCNCARWSNGAYACNW 922

Qy     626 taactgaaacagatgatgccagcaagccgcatgatgtttacataatgtcactggagt 682
             || ||:  : | || | :|       || |      |:| :| :  ::   |::
Db     923 SNACNGARGARCARGARGAYAARGCNGTNCARGTNWSNAAYAARAARAARMGNAARY 979


RESULT  11
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,974
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-974-2


  Query Match             1.7%;  Score 33.4;  DB 1;  Length 2277;
  Best Local Similarity   27.3%;  Pred. No. 0.82;
  Matches   81;  Conservative   52;  Mismatches  164;  Indels    0;  Gaps    0;

Qy     386 tgggagaggggaaagcattgatgaccgagaaaaaacagagcaggacagctatgaacctg 445
           ||| :||:| :||:| :  :|||:||:   ||  |    ||  |:||   :   |    :|
Db     683 TGGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGAYGAYGARGARGAYGGNGTNTTYG 742

Qy     446 ttgaagtgcaaggcctctcctacaaacgggacggcaaagggctgacgagaggtgcgtatg 505
           :|| | : | |         | ||    ||  : ||:|     | :   |   :  :
Db     743 AYGAYGARGAYGARGARGARGARAAYATHGARWSNAARGTNACNAARCCNGTNCARATHC 802

Qy     506 cccagcttgtgcttgtcagcgtaaagtcgggtgagatgaaagagctgacaagtcacaaag 565
             |::  |   |    :|        | |   |::  :  || |  ::     |  |
Db     803 ARAARMGNGCNGTNAARMGNCCNGCNCCNGCNAARWSNWSNGAYCAYWSNGARGARGAYW 862

Qy     566 ctgatcatggtgatcctgctttttctcctgacggcaaatggcttgttttctcagctaatt 625
           : ||::  |  ||  : | ::    :  :||:||  |:  ::| |     :: | :|  :
Db     863 SNGAYYTNGARGARWSNGAYWSNATHGAYGAYGGNGARGARYTNGCNCARWSNGAYACNW 922

Qy     626 taactgaaacagatgatgccagcaagccgcatgatgtttacataatgtcactggagt 682
             || ||:  : | || | :|       || |      |:| :| :  ::   |::
Db     923 SNACNGARGARCARGARGAYAARGCNGTNCARGTNWSNAAYAARAARAARMGNAARY 979


RESULT  12
US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, Kathleen
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,487
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-098-487-2
```

```
Query Match             1.7%;  Score 33.4;  DB 2;  Length 2277;
Best Local Similarity.  27.3%;  Pred. No. 0.82;
Matches   81;  Conservative   52;  Mismatches  164;  Indels    0;  Gaps    0;

Qy    386 tgggagaggggaaagcattgatgaccgagaaaaaacagagcaggacagctatgaacctg 445
          ||| :||:| :||:| :  :||:||:   ||  |    ||  |:||   :   |    :|
Db    683 TGGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGAYGARGARGAYGGNGTNTTYG 742

Qy    446 ttgaagtgcaaggcctctcctacaaacgggacggcaaagggctgacgagaggtgcgtatg 505
          :|| | : | |          | ||    ||  : ||:|     | :   |    :  :
Db    743 AYGAYGARGAYGARGARGARGARAAYATHGARWSNAARGTNACNAARCCNGTNCARATHC 802

Qy    506 cccagcttgtgcttgtcagcgtaaagtcgggtgagatgaaagagctgacaagtcacaaag 565
            |::  |   |    :|       | |   |::  :  || |  ::     |  |
Db    803 ARAARMGNGCNGTNAARMGNCCNGCNCCNGCNAARWSNWSNGAYCAYWSNGARGARGAYW 862

Qy    566 ctgatcatggtgatcctgctttttctcctgacggcaaatggcttgttttctcagctaatt 625
          : ||::  |  ||  : | ::    :  :||:||  |:  ::| |     :: | :|  :
Db    863 SNGAYYTNGARGARWSNGAYWSNATHGAYGAYGGNGARGARYTNGCNCARWSNGAYACNW 922

Qy    626 taactgaaacagatgatgccagcaagccgcatgatgtttacataatgtcactggagt 682
            || ||:  : | || | :|       || |      |:| :| :  ::   |::
Db    923 SNACNGARGARCARGARGAYAARGCNGTNCARGTNWSNAAYAARAARAARMGNAARY 979


RESULT  13
US-07-688-299-2
; Sequence 2, Application US/07688299
; Patent No. 5281525
;  GENERAL INFORMATION:
;    APPLICANT:  MITSUSHIMA, Kenji
;    APPLICANT:  TAKIMOTO, Akio
;    APPLICANT:  YAGI, Shigeo
;    APPLICANT:  SONOYAMA, Takayasu
;    TITLE OF INVENTION:  Cephalosporin Acetylhydrolase Gene and
;    TITLE OF INVENTION:  Protein Encoded by Said Gene
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  301 N. Washington Street, P. O. Box 747
;      CITY:  Falls Church
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/688,299
;      FILING DATE:  19910422
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-113483
;      FILING DATE:  27-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY, JR., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  20-2765P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 241-1300
;   INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  957 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..957
US-07-688-299-2


  Query Match             1.6%;  Score 31.8;  DB 1;  Length 957;
  Best Local Similarity   52.7%;  Pred. No. 1.5;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy   1489 gtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggc 1548
          | |  |||||  ||||   |         |  ||||    |   || |||||||| || ||
Db    478 GCGCTTGAGGTCATCAGCAGCTTCGACGAGGTTGACGAAACAAGGATCGGTGTGACAGGA 537

Qy   1549 ggaagctacggaggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagct 1608
          ||||||  | || ||    | | ||||         |  | || | |  ||      |||||
Db    538 GGAAGCCAAGGCGGAGGTTTAACCATTGCCGCAGCAGCGCTGTCAGACATTCCAAAAGCC 597

Qy   1609 gccgttaccca 1619
          || ||| || |
Db    598 GCGGTTGCCGA 608


RESULT  14
US-07-980-517A-2
; Sequence 2, Application US/07980517A
; Patent No. 5338676
;  GENERAL INFORMATION:
;    APPLICANT:  Kenji MITSUSHIMA
;    APPLICANT:  Akio TAKIMOTO
;    APPLICANT:  Shigeo YAGI
;    APPLICANT:  Takayasu SONOYAMA
;    TITLE OF INVENTION:  Cephalosporin Acetylhydrolase Gene
;    TITLE OF INVENTION:  and Protein Encoded by Said Gene
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5338676th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  United States of America
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/980,517A
;      FILING DATE:  19921123
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 113483/1990
;      FILING DATE:  27-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gerald M. Murphy, Jr.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  20-3177
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-241-1300
;      TELEFAX:  703-241-2848
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  957 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-07-980-517A-2


  Query Match             1.6%;  Score 31.8;  DB 1;  Length 957;
  Best Local Similarity   52.7%;  Pred. No. 1.5;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy   1489 gtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggc 1548
```

```
          | |  |||||  ||||   |        |  ||||    |   || |||||||  || ||
Db    478 GCGCTTGAGGTCATCAGCAGCTTCGACGAGGTTGACGAAACAAGGATCGGTGTGACAGGA 537

Qy   1549 ggaagctacggaggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagct 1608
          ||||||  | || ||   | | ||||       |  | || | |  ||      |||||
Db    538 GGAAGCCAAGGCGGAGGTTTAACCATTGCCGCAGCAGCGCTGTCAGACATTCCAAAAGCC 597

Qy   1609 gccgttaccca 1619
          || ||| || |
Db    598 GCGGTTGCCGA 608


RESULT  15
US-07-688-299-12
; Sequence 12, Application US/07688299
; Patent No. 5281525
;  GENERAL INFORMATION:
;    APPLICANT:  MITSUSHIMA, Kenji
;    APPLICANT:  TAKIMOTO, Akio
;    APPLICANT:  YAGI, Shigeo
;    APPLICANT:  SONOYAMA, Takayasu
;    TITLE OF INVENTION:  Cephalosporin Acetylhydrolase Gene and
;    TITLE OF INVENTION:  Protein Encoded by Said Gene
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  301 N. Washington Street, P. O. Box 747
;      CITY:  Falls Church
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/688,299
;      FILING DATE:  19910422
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-113483
;      FILING DATE:  27-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY, JR., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  20-2765P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 241-1300
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2046 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  184..1140
US-07-688-299-12


  Query Match             1.6%;  Score 31.8;  DB 1;  Length 2046;
  Best Local Similarity   52.7%;  Pred. No. 2.7;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy   1489 gtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggc 1548
          | |  |||||  ||||   |        |  ||||    |   || |||||||  || ||
Db    661 GCGCTTGAGGTCATCAGCAGCTTCGACGAGGTTGACGAAACAAGGATCGGTGTGACAGGA 720

Qy   1549 ggaagctacggaggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagct 1608
          ||||||  | || ||   | | ||||       |  | || | |  ||      |||||
Db    721 GGAAGCCAAGGCGGAGGTTTAACCATTGCCGCAGCAGCGCTGTCAGACATTCCAAAAGCC 780

Qy   1609 gccgttaccca 1619
          || ||| || |
Db    781 GCGGTTGCCGA 791


Search completed: September 26, 2001, 08:00:14
Job time: 17436 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2001, 17:34:53 ; Search time 34.68 Seconds
(without alignments)
1148.499 Million cell updates/sec

Title: US-09-462-845-2
Perfect score: 3489
Sequence: 1 MKKLITADDITAIVSVTDPQ.....HPRQRIKRLNYISSWFDQHL 657

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3489	100.0	657	20	AAY08553	B. subtilis hydrol
2	3489	100.0	657	20	AAW97789	Bacillus subtilis
3	1104.5	31.7	632	19	AAW70508	Pyrococcus horikos
4	625	17.9	622	19	AAW41248	Thermococcus amida
5	590.5	16.9	684	21	AAB18512	H3 homologue of pr
6	453	13.0	841	21	AAB18511	H2 homologue of pr
7	396.5	11.4	783	10	AAP94265	Sequence of APH36.
8	391	11.2	732	12	AAR11056	Acylamino acid-iso
9	383	11.0	732	13	AAR20001	Human acyl amino a
10	311	8.9	541	21	AAB18510	H1 homologue of pr
11	272	7.8	759	15	AAR54612	Delta3-9 CD26. Ho
12	272	7.8	766	14	AAR40909	Sequence encoded b
13	271	7.8	766	15	AAR54611	Native CD26. Homo
14	269	7.7	931	20	AAW89596	Saccharomyces cere
15	262.5	7.5	732	21	AAB18507	Amino acid sequenc
16	257.5	7.4	760	18	AAW27438	Human fibroblast a
17	247.5	7.1	739	15	AAR54613	Delta24-34 CD26.
18	244.5	7.0	767	21	AAB11748	Rat dipeptidyl pep
19	243	7.0	771	20	AAW89589	Aspergillus oryzae
20	238	6.8	771	20	AAW97798	Aspergillus oryzae
21	237.5	6.8	676	21	AAB18515	Amino acid sequenc
22	235	6.7	731	21	AAG38773	Arabidopsis thalia
23	232.5	6.7	623	21	AAG38774	Arabidopsis thalia
24	231.5	6.6	621	21	AAG28040	Arabidopsis thalia
25	223.5	6.4	569	21	AAG38775	Arabidopsis thalia
26	222.5	6.4	569	21	AAG28041	Arabidopsis thalia
27	219	6.3	759	18	AAW31963	Human fibroblast a
28	212.5	6.1	508	21	AAB42928	Human ORFX ORF2692
29	201.5	5.8	960	21	AAG48128	Arabidopsis thalia
30	201.5	5.8	968	21	AAG48127	Arabidopsis thalia
31	201.5	5.8	1063	21	AAG48126	Arabidopsis thalia
32	191.5	5.5	487	21	AAG28042	Arabidopsis thalia
33	184.5	5.3	710	21	AAB29603	Human prolyl endop
34	182.5	5.2	710	21	AAB29602	Porcine prolyl end
35	179.5	5.1	652	21	AAG29115	Arabidopsis thalia
36	179.5	5.1	657	21	AAG29114	Arabidopsis thalia
37	166.5	4.8	615	21	AAB29604	Pyrococcus furiosu
38	164.5	4.7	690	21	AAB29599	Aeromonas punctata
39	162.5	4.7	723	19	AAW40286	S. capsulata proly
40	160.5	4.6	668	22	AAB79081	Corynebacterium gl
41	160.5	4.6	690	21	AAB29600	Aeromonas hydrophi
42	156.5	4.5	351	21	AAG48134	Arabidopsis thalia
43	156.5	4.5	360	21	AAG48133	Arabidopsis thalia
44	156.5	4.5	381	21	AAG48132	Arabidopsis thalia
45	154	4.4	295	21	AAG48135	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAY08553
ID AAY08553 standard; protein; 657 AA.
XX
AC AAY08553;
XX
DT 03-AUG-1999 (first entry)
XX
DE B. subtilis hydrolase protein YUXL.
XX
KW Hydrolase; bacterial; Gram positive; YUXL; YTMA; YITV; TQKD; YCLE; YTAP;
KW YDEN; YBFK; YFHM; YDJP; YVFQ; YVAM; YQJL; SRFAD; YCGS; YTPA; YBAC; YUII;
KW YODD; YJCH; YODH; detection; cleaning; detergent; textile treatment;
KW animal feed; peptide hydrolysis; waste treatment; cleavage.
XX
OS Bacillus subtilis.
XX
PN WO9927081-A2.
XX
PD 03-JUN-1999.
XX
PF 19-NOV-1998; 98WO-US24973.
XX
PR 20-NOV-1997; 97GB-0024629.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Estell DA;
XX
DR WPI; 1999-347714/29.
XX
PT Hydrolases from Gram-positive microorganisms
XX
PS Claim 1; Page 20; 24pp; English.

```
XX
CC   This invention describes novel Gram positive bacterial hydrolases,
CC   namely YUXL, YTMA, YITV, TQKD, YCLE, YTAP, YDEN, YBFK, YFHM, YDJP, YVFQ,
CC   YVAM, YQJL, SRFAD, YCGS, YTPA, YBAC, YUII, YODD, YJCH and YODH. The
CC   encoding nucleic acid and its fragments, are useful as probes to detect
CC   Gram-positive hydrolases. The gram-positive microorganisms are used to
CC   produce recombinant hydrolases which are useful in cleaning (e.g.
CC   detergent) or textile treatment compositions, and in animal feeds. Other
CC   uses include peptide hydrolysis, waste treatment and cleavage of fusion
CC   proteins.
XX
SQ   Sequence   657 AA;

 Query Match             100.0%;  Score 3489;  DB 20;  Length 657;
 Best Local Similarity   100.0%;  Pred. No. 4.9e-286;
 Matches  657;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MKKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mkklitadditaivsvtdpqyapdgtraayvksqvnqekdsytsniwiyetktggsvpwt 60

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 hgekrstdprwspdgrtlafisdregdaaqlyimsteggearkltdipygvskplwspdg 120

Qy     121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 esilvtislgegesiddrekteqdsyepvevqglsykrdgkgltrgayaqlvlvsvksge 180

Qy     181 MKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGS 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 mkeltshkadhgdpafspdgkwlvfsanltetddaskphdvyimslesgdlkqvtphrgs 240

Qy     241 FGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 fgsssfspdgrylallgnekeyknatlskawlydieqgrltcltemldvhladaligdsl 300

Qy     301 IGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHF 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 iggaeqrpiwtkdsqgfyvigtdqgstgiyyisieglvypirlekeyinsfslspdeqhf 360

Qy     361 IASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMR 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 iasvtkpdrpselysiplgqeekqltgandkfvrehtisipeeiqyatedgvmvngwlmr 420

Qy     421 PAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 paqmegettyplilnihggphmmyghtyfhefqvlaakgyavvyinprgshgygqefvna 480

Qy     481 VRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 vrgdyggkdyddvmqavdeaikrdphidpkrlgvtggsyggfmtnwivgqtnrfkaavtq 540

Qy     541 RSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDD 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 rsisnwisfhgvsdigyfftdwqlehdmfedteklwdrsplkyaanvetpllilhgerdd 600

Qy     601 RCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 rcpieqaeqlfialkkmgketklvrfpnashnlsrtghprqrikrlnyisswfdqhl 657


RESULT   2
AAW97789
ID   AAW97789 standard; Protein; 657 AA.
XX
AC   AAW97789;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Bacillus subtilis serine protease SP1 (YUXL).
XX
KW   Serine protease SP1; YUXL; detergent.
XX
OS   Bacillus subtilis.
XX
PN   WO9903984-A2.
XX
PD   28-JAN-1999.
XX
PF   14-JUL-1998;   98WO-US14647.
XX
PR   15-JUL-1997;   97EP-0305232.
XX
PA   (GEMV ) GENENCOR INT BV.
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Estell DA;
XX
DR   WPI; 1999-132231/11.
DR   N-PSDB; AAX07301.
XX
PT   Use of serine protease genes from Gram-positive microorganisms - for
PT   modification of host cells for the production of heterologous
PT   proteins or for producing proteins for use in cleaning compositions
XX
PS   Disclosure; Fig 1A-C; 37pp; English.
XX
CC   This polypeptide comprises serine protease SP1 (YUXL) of Bacillus
CC   subtilis.  Novel serine proteases SP1, SP2, SP3, SP4 and SP5 (see
CC   AAW97789-93) were identified via a FASTA search of Bacillus subtilis
CC   genomic nucleic acid sequences.  SP1 was identified by its
CC   structural homology to the S9 type serine protease dap2 of yeast.
CC   SP2, SP3, SP4 and SP5 were identified by their structural and
CC   overall amino acid homology to SP1.  Host cells in which the
CC   naturally occurring gene encoding one or more of SP1, SP2, SP3,
CC   SP4 or SP5 (see AAX07301-04) is mutated such that the proteolytic
CC   activity is diminished or deleted, can be used for the production
CC   of heterologous proteins, e.g. a hormone, enzyme, growth factor,
CC   cytokine, protease, carbohydrase, lipase, racemase, epimerase,
CC   tautomerase, mutase, transferase, kinase or phosphatase (claimed).
CC   SP1, SP2, SP3, SP4 and SP5 can also be produced on a large scale
CC   in a microbial host expression system for use in cleaning
CC   compositions such as detergents, bar or liquid soap, dish-care
CC   formulations and contact lens cleaning solutions, or for peptide
CC   hydrolysis, waste treatment, textile applications, as
CC   fusion-cleavage enzymes in protein production, and as animal feed
CC   additives.
XX
SQ   Sequence   657 AA;

 Query Match             100.0%;  Score 3489;  DB 20;  Length 657;
 Best Local Similarity   100.0%;  Pred. No. 4.9e-286;
 Matches  657;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MKKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mkklitadditaivsvtdpqyapdgtraayvksqvnqekdsytsniwiyetktggsvpwt 60

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 hgekrstdprwspdgrtlafisdregdaaqlyimsteggearkltdipygvskplwspdg 120

Qy     121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 esilvtislgegesiddrekteqdsyepvevqglsykrdgkgltrgayaqlvlvsvksge 180

Qy     181 MKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGS 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 mkeltshkadhgdpafspdgkwlvfsanltetddaskphdvyimslesgdlkqvtphrgs 240

Qy     241 FGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 fgsssfspdgrylallgnekeyknatlskawlydieqgrltcltemldvhladaligdsl 300
```

```
Qy     301 IGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHF 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 iggaeqrpiwtkdsqgfyvigtdqgstgiyyisieglvypirlekeyinsfslspdeqhf 360

Qy     361 IASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMR 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 iasvtkpdrpselysiplgqeekqltgandkfvrehtisipeeiqyatedgvmvngwlmr 420

Qy     421 PAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 paqmegettyplilnihggphmmyghtyfhefqvlaakgyavvyinprgshgygqefvna 480

Qy     481 VRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 vrgdyggkdyddvmqavdeaikrdphidpkrlgvtggsyggfmtnwivgqtnrfkaavtq 540

Qy     541 RSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDD 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 rsisnwisfhgvsdigyfftdwqlehdmfedteklwdrsplkyaanvetpllilhgerdd 600

Qy     601 RCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 rcpieqaeqlfialkkmgketklvrfpnashnlsrtghprqrikrlnyisswfdqhl 657


RESULT   3
AAW70508
ID   AAW70508 standard; Protein; 632 AA.
XX
AC   AAW70508;
XX
DT   29-DEC-1998  (first entry)
XX
DE   Pyrococcus horikoshi acylpeptide hydrolase.
XX
KW   Pyrococcus horikoshi acylpeptide hydrolase.
XX
OS   Pyrococcus horikoshi.
XX
PN   JP10210977-A.
XX
PD   11-AUG-1998.
XX
PF   31-JAN-1997;   97JP-0018381.
XX
PR   31-JAN-1997;   97JP-0018381.
XX
PA   (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX
DR   WPI; 1998-488369/42.
DR   N-PSDB; AAV33582.
XX
PT   A heat-resistant acylpeptide hydrolase and a gene coding it - useful
PT   for hydrolysing the C termini of proteins at high temperature
XX
PS   Claim 2; Pages 5-6; 8pp; Japanese.
XX
CC   The present sequence represents a Pyrococcus horikoshi acylpeptide
CC   hydrolase.  The enzyme has the following properties: (a) it hydrolyses
CC   an acylpeptide; (b) an optimum temperature of 90-95 degrees Centigrade;
CC   (c) an optimum pH of 5.0-6.0; (d) no loss of activity when heated to 95
CC   degrees Centigrade for 3 hours at pH 7.5; and (e) a molecular weight of
CC   60 kDa.  The enzyme is claimed useful for hydrolysing the amino end of
CC   an acylated protein and a peptide at high temperatures.
XX
SQ   Sequence   632 AA;


 Query Match           31.7%;  Score 1104.5;  DB 19;  Length 632;
 Best Local Similarity 39.9%;  Pred. No. 1.3e-84;
 Matches  255;  Conservative  126;  Mismatches  217;  Indels   41;  Gaps   21;

Qy      27 RAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRTLAFIS--DR 84
           :| :  :::: : | | | :::|: |     |:| | | | :||:||:|: :|| |  |:
Db      25 kavfqvteislkdddyfsklylydgkr--vkpftsgnkds-nprfspngkliaftskrdk 81

Qy      85 EGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQD 144
           ||  ::||:: |:|||| |    ||:    :: ||:|| |   :        | ||   |
Db      82 egkeselyviptdggearllakfkygiknlrftedgksiavvtpi-------dvekkgnd 134

Qy     145 SYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLV 204
              :: : :  :| |   |    : || |:||: | ||    :     |  :|: |
Db     135 dvh--iireipfwfngvgwiygkrnvvylvdvesgkkkrltpknlnvdqirfh-ngr-ly 190

Qy     205 FSANLTETDDASKP--HDVYIMSLESGDLKQVTPHRGSFGSSSFSP--DGRYLALLGNEK 260
           |:|   : |   ||   |:|:  ||:  ::::||  | :    | |  || :: |  |
Db     191 fta---qedrerkplisdlyv--lenrkvrkltp--gkwrildflpldggsfv-lkantl 242

Qy     261 EYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVI 320
           |    | :  : || : | |   ||: || :  ::|  |  : |:::  :  |:   :||
Db     243 ergiptnahiyhydpktgelkkltkdldrnaynslnsd--vrgsqraelvykegwiyyv- 299

Qy     321 GTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIA--SVTKPDRPSELYSIPL 378
            ||     :: ::::| :  :       : || :  |   | |  :||    |:|||
Db     300 atdgpranlfrvnldgkiervigggdrsvesfdig-dyiaftaqdavt----ptelyiyrd 354

Qy     379 GQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHG 438
           | :||::|  | |::: :|:| ||  :    ||| :: |:|:|       || || |||
Db     355 g-kekkvtdfn-kwikgytlskpehfkvkasdgveidawvmkpvnfrkgkkypaileihg 412

Qy     439 GPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVD 498
           ||   ||: : ||| || :||: |:: ||||| |||:|| : :|| || :|| |:|: ||
Db     413 gpktaygyafmhefhvltskgfvvifsnprgsdgygeefad-irghygerdyqdlmevvd 471

Qy     499 EAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYF 558
           ||::|   || :|||||||||||||||||| ||||||||||||||||||| | :||||:
Db     472 ealrrfdfidgerlgvtggsyggfmtnwivghtnrfkaavtqrsisnwisffgttdigyy 531

Qy     559 FTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMG 618
           |   |:  | : : |  |::|||||| ||||||||:|   | || : :| ||||:|| :|
Db     532 fapdqigkdpwsnlegyweksplkyapnvetplliihstedyrcwlpealqlfislkylg 591

Qy     619 KETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
           |  :|  ||  :|:|||:| |: |:|||  |: | :: |
Db     592 krvelaifpgenhdlsrsgkpkhrvkrleliagwmekwl 630


RESULT   4
AAW41248
ID   AAW41248 standard; Protein; 622 AA.
XX
AC   AAW41248;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Thermococcus amidase.
XX
KW   Amidase; thermostable enzyme; optically active compound;
KW   L-amino acid; peptide; peptidomimetic; archaebacterium.
XX
OS   Thermococcus sp. strain GU5L5.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 410
FT                   /note= "encoded by TGC"
FT   Misc-difference 411
FT                   /note= "encoded by TGC"
XX
PN   WO9748794-A1.
XX
PD   24-DEC-1997.
XX
PF   17-JUN-1997;   97WO-US09319.
XX
PR   17-JUN-1996;   96US-0664646.
```

XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Murphy D, Reid JC, Robertson D;
XX
DR WPI; 1998-063135/06.
DR N-PSDB; AAV12887.
XX
PT Nucleic acid encoding heat-stable amidase from Thermococcus GU5L5 -
PT useful for removing N-terminal amino acids from synthetic peptide(s)
PT and peptidomimetic(s)
XX
PS Claim 1; Page 39-41; 56pp; English.
XX
CC This protein comprises a novel thermostable amidase of the
CC archaebacterium Thermococcus GU5L5. Its amino acid sequence was
CC deduced from an isolated nucleic acid (see AAV12887). Recombinant
CC amidase can be produced in transformed host cells, especially
CC prokaryotic cells. The amidase is used to remove Arg, Phe or Met
CC from the N-terminus of synthetic peptides or peptidomimetics
CC (claimed). Removal of the N-terminal residue can be done even in
CC the presence of a more reactive ester bond (very difficult to
CC achieve non-enzymatically). The amidase is selective for L-amino
CC acids and can therefore be used to produce optically active
CC compounds. The protein tolerates temperatures up to at least 70
CC degC and high concentrations of organic solvent (e.g. over 40%
CC dimethyl sulphoxide) so can cleave bonds that are normally
CC resistant.
XX
SQ Sequence 622 AA;

Query Match 17.9%; Score 625; DB 19; Length 622;
Best Local Similarity 29.1%; Pred. No. 3.9e-44;
Matches 192; Conservative 113; Mismatches 262; Indels 92; Gaps 27;

```
Qy      18 DPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRT 77
           ||:    |   ||  :: | : : | | : : :||         : :: || |||||
Db      19 dprir--gnliaytltkanmkdnkyestvvvedletgs----rrfienasmprispdgrk 72

Qy      78 LAFIS-DREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESID 136
           |||   : |    ::::   :  |:|:       |    |: |   :||      |
Db      73 laftcfneekketeiwvadiqtlsakkvlstk-nvrsmqwnddsrrllvv-----gfkrr 126

Qy     137 DREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAF 196
           | |    |   ||         |  |   |      ::  :: |: |          | |
Db     127 ddedfvfdddvpvwf-------dnmgffdgekttfwvldteaeeiie------qfekprf 173

Qy     197 SPDGKW----LVFSANLTETDDASKP-----HDVYIMSLESGDLKQVTPHRGSFGSSSFS 247
           |  | |    :| :    |    |||     :|: :   : |: ::::   | ||   :
Db     174 s-sglwhgdaivvnvphre---gskpalfkfydivlw--kdgeeeklf-ervsf--eavd 224

Qy     248 PDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTE-MLDVHLADALIGD--SLIGGA 304
           ||: : | | :|  |       ||| :  | |  : |  |||  |    |    |   |
Db     225 sdgkrillrgkkk--krfisehdwly-lwdgelkpiyegpldvweakltegkvyfltpda 281

Qy     305 EQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIASV 364
           :  :|  | :   |:      || ::|                 : |   :   :  :
Db     282 grvnlwlwdgkaervv------tgdhwi-----------------ygldvsdgkallli 317

Qy     365 TKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQM 424
               |  |||      | ||:|  |    |:      |   :: ::| : ::|| :||
Db     318 mtatrigelyl--ydgelkqvteyngpifrklktfeprhfrfkskd-leidgwylrpevk 374

Qy     425 EGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGD 484
           | :   |:|: :|||| |||| : :| |::|:|||   ::|||| || ::|   |
Db     375 eeka--pvivfvhggpkgmyghrfvyemqlmaskgyyccfvnprgsdgysedfalrvler 432

Qy     485 YGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSIS 544
           | :|::|:|  ::|  | :|  | :|:|:|| |||||||||| : |:: ||| :::  ||
Db     433 tgledfedimngieeffklepqadrervgitgisyggfmtnwaltqsdlfkagisengis 492

Qy     545 NWISFHGVSDIGYFFTDWQ------LEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
```

```
           |:: :  |||| :: | :      ||:: |    ||   ||| || ||: |:|::|
Db     493 ywltsyafsdiglwy-dvevigpnplenenf---rkl---splfyaqnvkapillihsle 545

Qy     599 DDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
           | |||::|:   :  || |||| :  |   :|  |  | || | ||     :|:: |
Db     546 dyrcpldqslmfynvlkdmgkeayiaifkrgahghsvrgsprhrpkryrlfiefferkl 604
```

RESULT 5
AAB18512
ID AAB18512 standard; protein; 684 AA.
XX
AC AAB18512;
XX
DT 15-JAN-2001 (first entry)
XX
DE H3 homologue of prolyl-tripeptidyl peptidase DPP.
XX
KW Prolyl tripeptidyl-peptidase; amidolytic activity; periodontal disease;
KW gingivitis; periodontitis.
XX
OS Porphyromonas gingivalis.
XX
PN WO200052147-A2.
XX
PD 08-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-US05551.
XX
PR 05-MAR-1999; 99US-0123148.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
PA (TRAV/) TRAVIS J.
PA (POTE/) POTEMPA J.
PA (BANB/) BANBULA A.
XX
PI Travis J, Potempa J, Banbula A;
XX
DR WPI; 2000-594181/56.
XX
PT Prolyl tripeptidyl-peptidase, active analog, fragment or variant useful
PT for identifying its inhibitor which is useful for protecting an animal
PT from a periodontal disease such as gingivitis and periodontitis -
XX
PS Claim 23; Fig 6; 58pp; English.
XX
CC The present sequence represents a H3 homologue of a prolyl
CC tripeptidyl-peptidase (designated DPP) from Porphyromonas gingivalis.
CC The prolyl tripeptidyl-peptidase has an amidolytic activity, and cleaves
CC a peptide bond in a target polypeptide having at least 4 amino acids.
CC This bond is between a proline and an amino acid attached to the
CC alpha-carboxyl group end of the proline. The polypeptide is useful
CC for identifying inhibitors. These inhibitors are then useful for
CC reducing the growth of bacterium or for protecting an animal from a
CC periodontal disease such as gingivitis and periodontitis caused by
CC Porphyromonas gingivalis.
XX
SQ Sequence 684 AA;

Query Match 16.9%; Score 590.5; DB 21; Length 684;
Best Local Similarity 25.8%; Pred. No. 3.8e-41;
Matches 179; Conservative 135; Mismatches 298; Indels 81; Gaps 23;

```
Qy       5 ITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEK 64
           :| :    :  |::   :|||  | |  |  : : :  |  ::       |   | |
Db      31 ltpelfmtlsrvsemalspdgktavyavsfpdvktnkatrelftvnldgsgrkqitdtes 90

Qy      65 RSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESIL 124
             | |  ||: :||:|: || : ||::|: :| | |:|::|  |::  |:||| : :|
Db      91 neyapawmadgkriafmsn-eggsmqlwvmnadgterrqlsnieggitgflfspdekqvl 149

Qy     125 VTISLGEGESIDDREKTEQDSYEPVE------VQGLSYKR----------------DGK 161
```

```
Db    150 ftkdikfg-------krtkdiypdldkatgriitdlmykhwdewvetiphpfianatdgm 202

Qy    162 GLTRGAYAQLVLVSVKSGEMKELTSHKADHG--DPAFSPDGKWLVFSANLTETDDA---S 216

Db    203 -ittgk-------dimegepye-apmkpwsgiedfswspdgqniay-asrkktgmaysls 252

Qy    217 KPHDVYIMSLESGDLKQVTPHRGSFGSSS---FSPDGRYLALLGNEKEYKNATLSKAWLY 273

Db    253 tnsdiyiynlasgrthnis--egmmgydtypkfspdgksiawismerdgyesdlkrlfva 310

Qy    274 DIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYIS 333

Db    311 dlatgkrthvnptfdynv------dmi--------qwapdskgiyflackeaetnlweit 356

Qy    334 IE-GLVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKF 392

Db    357 lktgkirqitqgqhdyadfsvrndv--mlakrhsfelpddlyrvnlkngaaqavtaenkv 414

Qy    393 VREHTISIPEEIQY-ATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGH--TYF 449

Db    415 ildrltpitcekrwmkttdggnmltwvvlppnfdknkkypailycqggpqntvsqfwsfr 474

Qy    450 HEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDP 509

Db    475 wnlrlmaeqgyiviapnrhgvpgfgqkwneqisgdyggqnmrdyltavde-mkkepyvdg 533

Qy    510 KRLGVTGGSYGGFMTNWIVG-QTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDM 568

Db    534 drigavgasyggfsvywlaghhdkrfaafiahagifn-lemqyatteemwfanwdiggpf 592

Qy    569 FED----TEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLV 624

Db    593 wekdnvvaqrtyatsphkfvqnwdtpilmihgeldfrilasqamaafdaaqlrgvpseml 652

Qy    625 RFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657

Db    653 iypdenhwvlq---pqnallfhrtffgwldrwl 682


RESULT   6
AAB18511
ID   AAB18511 standard; protein; 841 AA.
XX
AC   AAB18511;
XX
DT   15-JAN-2001  (first entry)
XX
DE   H2 homologue of prolyl-tripeptidyl peptidase DPP.
XX
KW   Prolyl tripeptidyl-peptidase; amidolytic activity; periodontal disease;
KW   gingivitis; periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200052147-A2.
XX
PD   08-SEP-2000.
XX
PF   03-MAR-2000; 2000WO-US05551.
XX
PR   05-MAR-1999;   99US-0123148.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
PA   (TRAV/) TRAVIS J.
PA   (POTE/) POTEMPA J.
PA   (BANB/) BANBULA A.
XX
PI   Travis J,  Potempa J,  Banbula A;
XX
DR   WPI; 2000-594181/56.
XX
PT   Prolyl tripeptidyl-peptidase, active analog, fragment or variant useful
PT   for identifying its inhibitor which is useful for protecting an animal
PT   from a periodontal disease such as gingivitis and periodontitis -
XX
PS   Claim 22; Fig 6; 58pp; English.
XX
CC   The present sequence represents a H2 homologue of a prolyl
CC   tripeptidyl-peptidase (designated DPP) from Porphyromonas gingivalis.
CC   The prolyl tripeptidyl-peptidase has an amidolytic activity, and cleaves
CC   a peptide bond in a target polypeptide having at least 4 amino acids.
CC   This bond is between a proline and an amino acid attached to the
CC   alpha-carboxyl group end of the proline. The polypeptide is useful
CC   for identifying inhibitors. These inhibitors are then useful for
CC   reducing the growth of bacterium or for protecting an animal from a
CC   periodontal disease such as gingivitis and periodontitis caused by
CC   Porphyromonas gingivalis.
XX
SQ   Sequence   841 AA;


  Query Match           13.0%;  Score 453;  DB 21;  Length 841;
  Best Local Similarity   23.5%;  Pred. No. 2.2e-29;
  Matches  174;  Conservative  121;  Mismatches  291;  Indels  154;  Gaps   29;

Qy      1 MKKLITADDITA--IVSVTDPQYAPD----------GT---------RA 28

Db    157 lksllllegdttatdvrvvlkpktardssalypnytgkerlsikhmmsgtflsggslsptg 216

Qy     29 AYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRTLAFISDREGDA 88

Db    217 kyvltsyrvsrdnkpavtynqlrdakgnillnlnekealg--wmphedmlmvir-kegna 273

Qy     89 AQLYIMSTEG-GEARKLTDIPYGVSKPLWSPDGESIL-VTISLGEGE--------SIDDR 138

Db    274 krlvafdpmgkgektlvsnlp--esqfrmspdaryylfykqekgpgkdplfirhldpddr 331

Qy    139 EKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSP 198

Db    332 qsdwrdr----------------------sqiyllnaesgvygpltfgysttyiydiap 368

Qy    199 DGKWLVFSANLTETDDASKPH---DVYIMSLESGDLKQVTPHRGSFGSSSFSPDGRYLAL 255

Db    369 dskraligt--lstdwtrrpfrfatimeynmetgkadtlitrdpsidaiqytpdgkhliv 426

Qy    256 LGNEKEYKNATLS-----------KAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGG 303

Db    427 mgsadafgniglnlksgvtpnsydkqfflfdlstrkataltknfn---------psvsag 477

Qy    304 AEQRPIWTKDSQGFYVIGTDQGS-TGIYYISIEGL-VYPIRLEKEYINSFSLSPDEQHFI 361

Db    478 rfdr------knnyyyfraengsrkqlyrldlktleisqiqtgedvvqwfgvaadngavw 531

Qy    362 ASVTKPDRPSELYSIPLGQEEKQLTGANDKFV----REHTISI----PEEIQYATEDGVM 413

Db    532 ysgqsannadrly---------rldgtkgklvwdlsaeklanidftpardwnytapdgtv 582

Qy    414 VNGWLMRPAQMEGETTYPLILNIHGGP----HMMYGHTYFHEFQVLAAKGYAVVYINPRG 469

Db    583 vegwyylppqfdpskkypmlvyyyggtspinrtlegh---yslamyaaqgyvvytlnpsg 639

Qy    470 SHGYGQEF----VNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTN 525

Db    640 ttgygqeyaarhvna----wgdrtadeiigatkefirthsfvngkkvgcfgasyggfmtq 695

Qy    526 WIVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKL-WD------- 577

Db    696 ylqtktdifaaavshagissisnywg---sgy----wgmgystvastdsypwnnpdlyag 748

Qy    578 RSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTG 637

Db    749 hsplfradkihtplllhgsvdtnvptaesvnlynalkilgrevefiefteqdhfil--- 805

Qy    638 HPRQRIKRLNYISSWFDQHL 657
```

```
Db    806 eperrirwtnsicawfarwl 825


RESULT   7
AAP94265
ID   AAP94265 standard; protein; 783 AA.
XX
AC   AAP94265;
XX
DT   24-JUN-1990  (first entry)
XX
DE   Sequence of APH36.1 clone.
XX
KW   Clone APH36.1; acyl-peptide hydrolase.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..721
FT                   /label=claimed protein
FT   Active-site     580..587
FT   Region          130.
FT                   /note="potential site of glycosylation"
FT   Region          131.
FT                   /note="potential site of glycosylation"
FT   Region          132.
FT                   /note="potential site of glycosylation"
FT   Region          229.
FT                   /note="potential site of glycosylation"
FT   Region          230.
FT                   /note="potential site of glycosylation"
FT   Region          231.
FT                   /note="potential site of glycosylation"
FT   Region          239.
FT                   /note="potential site of glycosylation"
FT   Region          240.
FT                   /note="potential site of glycosylation"
FT   Region          241.
FT                   /note="potential site of glycosylation"
XX
PN   EP303997-A.
XX
PD   22-FEB-1989.
XX
PF   13-AUG-1988;   88EP-0113186.
XX
PR   21-AUG-1987;   87US-0087936.
XX
PA   (GEHO-) THE GENERAL HOSPITAL CORP.
XX
PI   Smith JA;
XX
DR   WPI; 1989-055444/08.
DR   N-PSDB; AAN91052.
XX
PT   Recombinant acyl-peptide hydrolase -used to catalyse hydrolysis of N-acyl
PT   peptide(s) or reaction of N-acetyl amino acid donor and acceptor protein.
XX
PS   Disclosure; ; 21pp; English.
XX
CC   Sequence contains the claimed sequence of acyl-peptide hydrolase (APH)
CC   (claim 1, page 11).  APH can be used to catalyse the hydrolysis of the
CC   N-terminal acyl amino acid of an acylated polypeptide, or the reaction
CC   between a derivatised N alpha-acetyl amino acid donor and acceptor with
CC   a free alpha-NH2.  It can also be used to make refractory proteins
CC   susceptible to Edman sequencing or th reduce degradation of proteins to
CC   be used therapeutically.
XX
SQ   Sequence   783 AA;


 Query Match             11.4%;  Score 396.5;  DB 10;  Length 783;
 Best Local Similarity   21.7%;  Pred. No. 1.1e-24;

  Matches  148;  Conservative  126;  Mismatches  271;  Indels  137;  Gaps   21;

Qy     22 APDGTRAAYVK---SQVNQEKDSYTSNIW------------------IYETKTGGSVPW 59
          :| ||  | ::     |: |:  :   :|                    :||    | : |
Db     93 spsgtmkavlrkaggtvsgeekqfl-evweknrklksfnlsalekhgpvyeddcfgclsw 151

Qy     60 THGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPD 119
          :| |              | ::::::                                |
Db    152 shseth-----------llyvaekk-----------------------------rpk 168

Qy    120 GESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSG 179
           ||   | :|    | |:  : ::        | : |:  |: :   :   | :: : ||
Db    169 aesffqtkaldisasddemarpkkpdqaikgdqfvfyedwgetmvsksipvlcvldidsg 228

Qy    180 EMKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDL 231
           : |         |   ::|    :||         |      ::   :| : |  |
Db    229 nisvlegvpenvspgqafwapgdtgvvfvgwwhepfrlgiryctnrrsalyyvdlsggkc 288

Qy    232 KQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHL 291
           : ::    : |   |||   : |       :   |:  |||      | :| :: | :
Db    289 ellsdgslaicsprlspdqcrivylqypclaphhqcsqlclyd----wytkvtsvv-vdi 343

Qy    292 ADALIGDSLIG-GAEQRPI--WTKDS-----------QGFYVIGTDQGSTGIYYISIEG 336
              :|:|  |      |:  |: ||             |  : : |  || |  ::  |
Db    344 vprqlgesfsgiycsllplgcwsadsqrvvfdsaqrsrqdlfavdtqtgs--itsltaag 401

Qy    337 LVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSEL---YSIPLGQEE----------K 383
           :|         |: |:   :|  : |  |  |    :  | |:|:          :
Db    402 sagswkl---------ltidkdlmvaqfstpslppslkvgflpppgkeqsvswvsleeae 452

Qy    384 QLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMM 443
           : | :      |     | :|||  |   :   |::|:    :|  |:::  |||||
Db    453 pipgihwgvrvlhpppdqenvqyadldfeai---llqpsnppdktqvpmvvmphggphss 509

Qy    444 YGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKR 503
           :   :     :|   |:||: :| ||| |:||: : :: |: | :|  ||  ||:: ::
Db    510 fvtawmlfpamlckmgfavllvnyrgstgfgqdsilslpgnvghqdvkdvqfaveqvlqe 569

Qy    504 DPHIDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDW 562
           : | | :|: : |||:|||::  ::||    : | : :  : | |  | :||     ||
Db    570 e-hfdarrvalmggshggflschligqypetysaciarnpviniasmmgstdi----pdw 624

Qy    563 QLEHDMFEDT----------EKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFI 612
            :    |  :          |:: |:||:||   |:||:|:: |: | | | :|  : :
Db    625 cmvetgfpysnsclpdlnvweemldkspikyipqvktpvllmlgqedrrvpfkqgmeyyr 684

Qy    613 ALKKMGKETKLVRFPNASHNLS 634
           |||      :|: :| ::| ||
Db    685 alkarnvpvrlllypksnhals 706


RESULT   8
AAR11056
ID   AAR11056 standard; Protein; 732 AA.
XX
AC   AAR11056;
XX
DT   23-MAY-1991  (first entry)
XX
DE   Acylamino acid-isolating enzyme-like polypeptide.
XX
KW   AARE; pig liver.
XX
OS   Sus scrofa.
XX
PN   JP03030673-A.
XX
PD   08-FEB-1991.
XX
PF   29-JUN-1989;   89JP-0165216.
XX
PR   29-JUN-1989;   89JP-0165216.
XX
```

```
PA   (TAKA-) TAKARA SHUZO KK.
XX
DR   WPI; 1991-084340/12.
DR   N-PSDB; AAQ10958.
XX
PT   Acylamino acid-isolating enzyme-like polypeptide - prepd. by
PT   genetic engineering.
XX
PS   Claim 1; Fig 2; 10pp; Japanese.
XX
CC   The sequence was obtd. from six clones isolated from a pig liver
CC   cDNA library.  The clones, lambda AARE 419, 450, 451, 452, 521 and
CC   522 cover the entire sequence of the gene.
XX
SQ   Sequence   732 AA;
```

```
  Query Match             11.2%;  Score 391;  DB 12;  Length 732;
  Best Local Similarity   21.0%;  Pred. No. 3e-24;
  Matches  149;  Conservative  123;  Mismatches  287;  Indels  152;  Gaps   23;

Qy      22 APDGTRAAYVKSQ--VNQEKDSYTSNIW----------------IYETKTGGSVPWT 60
Db      97 spsgtmkavlrkaggtgtaeekqflevweknrklrsfnlsalekhgpvyeddcfgclsws 156

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
Db     157 hseth-----------llyvaekk----------------------------rpka 173

Qy     121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180
Db     174 esffqtkaldvtgsddemartkkpdqaikgdqflfyedwgenmvskstpvlcvldiesgn 233

Qy     181 MKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDLK 232
Db     234 isvlegvpesvspgqafwapgdtgvvfvgwwhepfrlgirfctnrrsalyyvdltggkce 293

Qy     233 QVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLA 292
Db     294 llsdesvavtsprlspdqcrivylrfpslvphqqcgqlclyd----wytrvtsvv-vdiv 348

Qy     293 DALIGDSLIG-GAEQRPI--WTKDS-----------QGFYVIGTDQGSTGIYYISIEGL 337
Db     349 prqlgedfsgiycsllplgcwsadsqrvvfdspqrsrqdlfavdtqmgsvtsltaggsgg 408

Qy     338 VYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSEL---YSIPLGQEEKQLTGANDKFVR 394
Db     409 swkl----------ltidrdlmvvqfstpsvppslkvgflppagkeq-----avswvsl 452

Qy     395 EHTISIP--------------EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGG 439
Db     453 eeaepfpdiswsirvlqpppqqehvqya---gldfeaillqpsnspektqvpmvvmphgg 509

Qy     440 PHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDE 499
Db     510 phssfvtawmlfpamlckmgfavllvnyrgstgfgqdsilslpgnvghqdvkdvqfaveq 569

Qy     500 AIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYF 558
Db     570 vlqee-hfdagrvalmggshggflschligqypetysacvvrnpviniasimgstdi--- 625

Qy     559 FTDWQLEHDMFEDTE------KLW----DRSPLKYAANVETPLLILHGERDDRCPIEQAE 608
Db     626 -pdwcmveagfsyssdclpdlsvwaamldkspikyapqvktpllmlgqedrrvpfkqgm 684

Qy     609 QLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISS--WFDQHL 657
Db     685 eyyrvlkarnvpvrlllypksthalsev-----evesdsfmnavlwlcthl 730
```

```
RESULT   9
AAR20001
ID   AAR20001 standard; Protein; 732 AA.
XX
AC   AAR20001;
XX
DT   27-MAR-1992  (first entry)
XX
DE   Human acyl amino acid isolating enzyme-like polypeptide.
XX
KW   recombinant technique.
XX
OS   Homo sapiens.
XX
PN   JP03254680-A.
XX
PD   13-NOV-1991.
XX
PF   20-JUL-1990;   90JP-0190584.
XX
PR   23-JAN-1990;   90JP-0011718.
PR   20-JUL-1990;   90JP-0190584.
XX
PA   (TAKA-) TAKARA SHUZO KK.
XX
DR   WPI; 1992-002453/01.
DR   N-PSDB; AAQ20001.
XX
PT   Human acylamino acid-isolating enzyme-like polypeptide - obtd. by
PT   culturing transformant contg. introduced recombinant plasmid
PT   contg. DNA coding for polypeptide
XX
PS   Claim 1; Fig 1; 14pp; Japanese.
XX
CC   A human liver cell cDNA library was screened with a cDNA fragment of
CC   porcine acyl amino acid acid isolating enzyme-like polypeptide as
CC   probe. A positive clone was analysed and sequenced.
XX
SQ   Sequence   732 AA;
```

```
  Query Match             11.0%;  Score 383;  DB 13;  Length 732;
  Best Local Similarity   21.0%;  Pred. No. 1.4e-23;
  Matches  147;  Conservative  136;  Mismatches  287;  Indels  130;  Gaps   24;

Qy      22 APDGTRAAYVKS---------------QVNQEKDSYTSNIW-----IYETKTGGSVPWT 60
Db      97 spsgsmkavlrkaggtgpgeekqflevweknrklksfnlsvlekhgpvyeddcfgclsws 156

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
Db     157 hseth-----------llyvaerk----------------------------rpka 173

Qy     121 ESILVTISLGEGESIDDREKTEQDSYEPVE-VQGLSYKRDGKGLTRGAYAQLVLVSVKSG 179
Db     174 esffqtkaldvsasddeiarlkkpd-qpikgdqfvfyedwgenmvsksipvlcvldvesg 232

Qy     180 EMKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDL 231
Db     233 nisvlegvpenvspgqafwapgdagvvfvgwwhepfrlgirfctnrrsalyyvdliggkc 292

Qy     232 KQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHL 291
Db     293 ellsddslavssprlspdqcrivylgypsliphhqcsqlclyd----wytkvtsvv-vdv 347

Qy     292 ADALIGDSLIG-GAEQRPI--WTKDSQGFYVIGTDQGSTGIYYISIE-GLVYPIRLEKEY 347
Db     348 vprqlgenfsgiycsllplgcwsadsqrvvfdsaqrsrqdlfavdtqvgtvtsltaggsg 407

Qy     348 INSFSLSPDEQHFIASVTKPDRPSELY--SIPLGQEEKQLTGANDKFVREHTISIP---- 401
Db     408 gswkllticdqdlmvaqfstpslpptlkvgflpsagkeqsvlwvs----leeaepipdihw 463

Qy     402 -----------EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFH 450
Db     464 girvlqpppeqenvqyarldfeai---llqpgsppdktqvpmvvmphggphssfvtawml 520
```

```
Qy     451 EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPK 510
Db     521 fpamlckmgfavllvnyrgstgfgqdsilslpgnvghqdvkdvqfaveqvlqee-hfdas 579

Qy     511 RLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDW------- 562
Db     580 hvalmggshggfischligqypetyracvarnpviniasmlgstdi----pdwcvveagf 635

Qy     563 QLEHDMFEDTE---KLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGK 619
Db     636 pfssdclpdlsvwaemldkspiryipqvktplllmlgqedrrvpfkqgmeyyralktrnv 695

Qy     620 ETKLVRFPNASHNLSRTGHPRQRIKRLNYISS--WFDQHL 657
Db     696 pvrlllypksthalsev-----evesdsfmnavlwlrthl 730
```

RESULT 10
AAB18510
ID AAB18510 standard; protein; 541 AA.
XX
AC AAB18510;
XX
DT 15-JAN-2001 (first entry)
XX
DE H1 homologue of prolyl-tripeptidyl peptidase DPP.
XX
KW Prolyl tripeptidyl-peptidase; amidolytic activity; periodontal disease;
KW gingivitis; periodontitis.
XX
OS Porphyromonas gingivalis.
XX
PN WO200052147-A2.
XX
PD 08-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-US05551.
XX
PR 05-MAR-1999; 99US-0123148.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
PA (TRAV/) TRAVIS J.
PA (POTE/) POTEMPA J.
PA (BANB/) BANBULA A.
XX
PI Travis J, Potempa J, Banbula A;
XX
DR WPI; 2000-594181/56.
XX
PT Prolyl tripeptidyl-peptidase, active analog, fragment or variant useful
PT for identifying its inhibitor which is useful for protecting an animal
PT from a periodontal disease such as gingivitis and periodontitis -
XX
PS Claim 21; Fig 6; 58pp; English.
XX
CC The present sequence represents a H1 homologue of a prolyl
CC tripeptidyl-peptidase (designated DPP) from Porphyromonas gingivalis.
CC The prolyl tripeptidyl-peptidase has an amidolytic activity, and cleaves
CC a peptide bond in a target polypeptide having at least 4 amino acids.
CC This bond is between a proline and an amino acid attached to the
CC alpha-carboxyl group end of the proline. The polypeptide is useful
CC for identifying inhibitors. These inhibitors are then useful for
CC reducing the growth of bacterium or for protecting an animal from a
CC periodontal disease such as gingivitis and periodontitis caused by
CC Porphyromonas gingivalis.
XX
SQ Sequence 541 AA;

Query Match 8.9%; Score 311; DB 21; Length 541;
Best Local Similarity 22.9%; Pred. No. 1.1e-17;
Matches 143; Conservative 105; Mismatches 223; Indels 154; Gaps 29;

```
Qy      91 LYIMSTEGGEARKLTDIPYGVSKPLWSPDG--ESILVTISLGEGESIDDREKTEQDSYEP 148
Db      11 lfasnidgsntrdlt--pf---------dgvkasilnmlkeqkdymiismnknnpqifep 59

Qy     149 VE-------------------VQGLSYKRDGK-----GLTRGAYAQLVLVSVKSGEMKE 183
Db      60 yklnvvtgeltqlyenkdaanpiqgyefdkdgelrgysrlvngieselyykdlatgefrl 119

Qy     184 LTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHD-VYIM-SLESG-------DLKQV 234
Db     120 l---kkthwddtfg------viafny-----asknkdeayvltnldsdktrivlydlkqn 165

Qy     235 TPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIE----QGRLTCLTEMLDVH 290
Db     166 kiireifanedydvsglhlsrkrn-------------yeidlmayegeksvvvpvsaty 211

Qy     291 LADALIGDSLIGGAEQRPIWTKDSQG--FYVIGTDQGSTGIYYISIEGLVYPIRLEKEYI 348
Db     212 -------------kelhklmekefkgkefsvvdydddet-illiavqs-------dklyg 250

Qy     349 NSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYAT 408
Db     251 tyyqfd----------trtkkftllydl-mpq------------lkeedmaemrpikfks 287

Qy     409 EDGVMVNGWLMRP-AQMEGETTYPLILNIHGGPHMMYGHTYFH-EFQVLAAKGYAVVYIN 466
Db     288 rdgltihgfitlpkaalegkkv-plivnphggpqgirdswgfnpetqlfasrgyatlqvn 346

Qy     467 PRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNW 526
Db     347 frisggygkeflragfkqigrkamddvedgvryaisqg-wvdpdriaiygashggyat-- 403

Qy     527 IVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFED-TEKLW--------- 576
Db     404 lmglvk------tpdlyacgvdyvgvsniytffdsfpeywkpfkemvkeiwydldnpeea 457

Qy     577 ----DRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHN 632
Db     458 aiakevspffqidkinkplfvvqgandprvninesdqivtalrargfevpymvkyneghg 517

Qy     633 LSRTGHPRQRIKRLNYISSWFDQHL 657
Db     518 fhreensmelyram---lgffakhl 539
```

RESULT 11
AAR54612
ID AAR54612 standard; Protein; 759 AA.
XX
AC AAR54612;
XX
DT 09-DEC-1994 (first entry)
XX
DE Delta3-9 CD26.
XX
KW Human; T cell activation antigen; CD26; analogues; deletion; soluble;
KW signal peptidase; immune-stimulating; response-stimulating; AIDS;
KW immunosuppression; AIDS-related complex.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 2..3
FT /note= "Position of delta3-9 deletion"
XX
PN WO9409132-A.
XX
PD 28-APR-1994.
XX
PF 19-AUG-1993; 93WO-US07923.
XX
PR 21-AUG-1992; 92US-0934162.

```
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
XX
PI   Morimoto C,  Schlossman S,  Tanaka T;
XX
DR   WPI; 1994-151317/18.
XX
PT   Polypeptide fragments and analogues of CD26 and encoding nucleic
PT   acid - useful for stimulating immune response, e.g. for treatment
PT   of AIDS to counteract immunosuppressive drug, and as vaccine
PT   adjuvant.
XX
PS   Claim 3; Page 49-52; 85pp; English.
XX
CC   The sequences given in AAR54612-14 represents analogues of the human T
CC   cell activation antigen CD26 which have internal deletions.  The
CC   analogues pref. lack residues 3-9 or 24-34.  These analogues are
CC   soluble under physiological conditions and lack enough amino acid
CC   residues to render them susceptible to cleavage by signal peptidase.
CC   The peptide fragments and analogues are useful as immune or response-
CC   stimulating therapeutics, eg. they may be used for treatment of
CC   disease conditions characterised by immunosuppression, eg. AIDS or
CC   AIDS-related complex, other virally or environmentally-induced
CC   conditions, and certain congenital immune deficiencies.  The peptides
CC   can be employed to increase immune function which has been impaired by
CC   use of immunosupressive drugs, such as certain chemotherapeutic drugs.
XX
SQ   Sequence   759 AA;


  Query Match             7.8%;  Score 272;  DB 15;  Length 759;
  Best Local Similarity   20.4%;  Pred. No. 3.6e-14;
  Matches  153;  Conservative  113;  Mismatches  285;  Indels  200;  Gaps   34;

Qy      15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
           |: |   :|||        : | | : |||::  ||         | :    :  |
Db      94 sindysispdgqfilleynyvkqwrhsytasydiydlnkrqliteeripnntqwv----- 148

Qy      66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
             ||| |  ||:: : :     :|:       : ::|     || |
Db     149 ----twspvghklayvwnnd-----iyvkiepnlpsyritwtgkediiyngitdwvyeee 199

Qy     110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
               |   |||:|  :             :| |    |   |    :|
Db     200 vfsaysalwwspngtflaya-------qfndtevplieysfysdeslq------------ 240

Qy     165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
           | : | |   |:|               | :|  |: |        ||  |
Db     241 ---ypktvrvpypkag--------------avnptvkffv-----vntdsls------- 270

Qy     224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
           |: :    |:|        :|      ||        :   ||   :|  ||
Db     271 -svtnatsiqit------apasmligdhylc------dvtwatqerislqwlrriqnysv 317

Qy     273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
               ||   ||  ||      : ::          :|       | :| |   || :|  ::|
Db     318 mdicdydessgrwnclvarqhiemstt----gwvgrfrpsephftldgnsfykiisneeg 373

Qy     326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
           | |  |:     :  :   | |   :|: |  ::|::  |
Db     374 yrhicyfqidkkdctfitkgtwevigiealtsdylyyisneykgmpggrnlykiqlsdyt 433

Qy     367 ----------PDRPSELYSIPLGQEEK--QL-------------TGANDKFVR--EHTIS 399
                   |:|  : ||:   :| |  ||             :  ||| :|  |   :
Db     434 kvtclscelnper-cqyysvsfskeakyyqlrcsgpglplytlhssvndkglrvlednsa 492

Qy     400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
           : : :|         | :::|     |  :: |    :     |||:|::: ||      | |
Db     493 ldkmlqnvqmpskkldfiilnetkfwyqmilpphfdkskkyplllдvyagpcsqkadtvf 552

Qy     450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
               : : :   |   : |||   | : ::|:    |  : :| ::|  :  |
Db     553 rlnwatylasteniivasfdgrgsgyqgdkimhainrrlgtfevedqieaarqfskmg-f 611

Qy     507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
           :| ||: : | ||||::|: ::|  :  ||  :    :| |  :  |      :|:  :
Db     612 vdnkriaiwgwsyggyvtsmvlgsgsgvfkcgiavapvsrweyydsv------yterymg 665

Qy     566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
             ||    :  | :   |    :   |::||  ||    :|: |:  ||  :| : :
Db     666 lptpednldhyrnstvmsraenfkqveyllihgtaddnvhfqqsaqiskalvdvgvdfqa 725

Qy     624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653
           : : :  | :| | |         ::|   |
Db     726 mwytdedhgiasstahqhiythmshfikqcf 756


RESULT  12
AAR40909
ID   AAR40909 standard; Protein; 766 AA.
XX
AC   AAR40909;
XX
DT   05-FEB-1994  (first entry)
XX
DE   Sequence encoded by human CD26 cDNA.
XX
KW   Human T cell activation antigen; monoclonal antibody Ta1.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          7..28
FT                   /label= hydrophobic
FT   Region          29..323
FT                   /label= N-terminal glycosylated region of
FT                         extracellular domain
FT                   /note= "8 sites for N-linked glycans"
FT   Region          324..551
FT                   /label= Cysteine rich region of extracellular
FT                         domain
FT                   /note= "1 N-linked glycosylation site"
FT   Region          552..766
FT                   /label= C-terminal region of extracellular
FT                         domain
FT                   /note= "1 N-linked glycosylation site & 1
FT                         catalytic site"
FT   Active-site     627..631
FT                   /label= active site of serine protease/esterase
FT                   /note= "fits the consensus sequence GXSXG"
XX
PN   WO9316102-A.
XX
PD   19-AUG-1993.
XX
PF   09-APR-1992;   92WO-US02892.
XX
PR   06-FEB-1992;   92US-0832211.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
XX
PI   Morimoto C,  Schlossman SF,  Tanaka T;
XX
DR   WPI; 1993-272827/34.
DR   N-PSDB; AAQ46089.
XX
PT   Polypeptide fragments of CD26 - are capable of disrupting binding
PT   of CD45 and CD26 and thus interfering with T-cell activation
XX
PS   Disclosure; pages 39-43; 73pp; English.
XX
CC   C26 is a human T cell antivation antigen originally identified by
CC   its reactivity with the MAb Ta1. C26 cDNA library was constructed
CC   from human PHA-activated T cells using the CDM7vector.The hydrophobic
CC   N-terminal of the predicted CD26 polypeptide has the characteristics
CC   of a signal sequence of the type II membrane protein, which is
```

CC reinforced by the observation that potential N-glycosylation sites
CC are located in the carboxy side of the hydrophobic core.Therefore
CC the N-terminal 6 AAs are predicted to be cytoplasmic, the next 22
CC AAs are predicted to transverse the cytoplasmic membrane, and the
CC 738 C-terminal AAs constitute the predicted extracellular domain.
XX
SQ Sequence 766 AA;

Query Match 7.8%; Score 272; DB 14; Length 766;
Best Local Similarity 20.4%; Pred. No. 3.6e-14;
Matches 153; Conservative 113; Mismatches 285; Indels 200; Gaps 34;

```
Qy      15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
           |: |   :|||       : | | : |||::  ||         | :   :  |
Db     101 sindysispdgqfilleynyvkqwrhsytasydiydlnkrqliteeripnntqwv----- 155

Qy      66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
             ||| |  ||:: : :     :|:      : ::|     || |
Db     156 ----twspvghklayvwnnd-----iyvkiepnlpsyritwtgkediiyngitdwvyeee 206

Qy     110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
               |   |||:|  :           :| |    |   |    :|
Db     207 vfsaysalwwspngtflaya-------qfndtevplieysfysdeslq----------- 247

Qy     165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
             | : | |   |:|               | :|  |: |       ||  |
Db     248 ---ypktvrvpypkag--------------avnptvkffv-----vntdsls------- 277

Qy     224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
            |: :    |:|        :|      ||       :   ||   :|  ||
Db     278 -svtnatsiqit------apasmligdhylc------dvtwatqerislqwlrriqnysv 324

Qy     273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
                 ||   ||  ||     : ::        :|       | :| |   || :|  ::|
Db     325 mdicdydessgrwnclvarqhiemstt----gwvgrfrpsephftldgnsfykiisneeg 380

Qy     326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
             | |  |:    :  :   | |   :|: |  ::|::  |
Db     381 yrhicyfqidkkdctfitkgtwevigiealtsdylyyisneykgmpggrnlykiqlsdyt 440

Qy     367 ----------PDRPSELYSIPLGQEEK--QL-----------TGANDKFVR--EHTIS 399
                    |:|  : ||:   :| |  ||            :  ||| :|  |   :
Db     441 kvtclscelnper-cqyysvsfskeakyyqlrcsgpglplytlhssvndkglrvlednsa 499

Qy     400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
           : : :|        | :::|    |  :: |   :    |||:|::: |\|      | |
Db     500 ldkmlqnvqmpskkldfiilnetkfwyqmilpphfdkskkyplldvyagpcsqkadtvf 559

Qy     450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
               : :   |   : |||   | : ::|:    | : :| ::|  :  |
Db     560 rlnwatylasteniivasfdgrgsgyqggdkimhainrrlgtfevedqieaarqfskmg-f 618

Qy     507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
           :| ||: : | ||||::|: ::|  :  ||  :    :| |  :  |      :|:  :
Db     619 vdnkriaiwgwsyggyvtsmvlgsgsgvfkcgiavapvsrweyydsv------yterymg 672

Qy     566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
             ||    :  | :   |    :   |::||  ||    :|: |:  ||  :| : :
Db     673 lptpednldhyrnstvmsraenfkqveyllihgtaddnvhfqqsaqiskalvdvgvdfqa 732

Qy     624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653
           : : :  | : | | |        ::|   |
Db     733 mwytdedhgiasstahqhiythmshfikqcf 763
```

RESULT 13
AAR54611
ID AAR54611 standard; Protein; 766 AA.
XX
AC AAR54611;
XX
DT 09-DEC-1994 (first entry)

XX
DE Native CD26.
XX
KW Human; T cell activation antigen; CD26; analogues; deletion; soluble;
KW signal peptidase; immune-stimulating; response-stimulating; AIDS;
KW immunosuppression; AIDS-related complex.
XX
OS Homo sapiens.
XX
PN WO9409132-A.
XX
PD 28-APR-1994.
XX
PF 19-AUG-1993; 93WO-US07923.
XX
PR 21-AUG-1992; 92US-0934162.
XX
PA (DAND) DANA FARBER CANCER INST INC.
XX
PI Morimoto C, Schlossman S, Tanaka T;
XX
DR WPI; 1994-151317/18.
DR N-PSDB; AAQ63261.
XX
PT Polypeptide fragments and analogues of CD26 and encoding nucleic
PT acid - useful for stimulating immune response, e.g. for treatment
PT of AIDS to counteract immunosuppressive drug, and as vaccine
PT adjuvant.
XX
PS Disclosure; Page 46-49; 85pp; English.
XX
CC This sequence represents the human T cell activation antigen CD26.
CC The invention is concerned with polypeptide fragments and analogues
CC of CD26 which have internal deletions (see also AAR54612-14). The
CC analogues pref. lack residues 3-9 or 24-34. These analogues are
CC soluble under physiological conditions and lack enough amino acid
CC residues to render them susceptible to cleavage by signal peptidase.
CC The peptide fragments and analogues are useful as immune or response-
CC stimulating therapeutics, eg. they may be used for treatment of
CC disease conditions characterised by immunosuppression, eg. AIDS or
CC AIDS-related complex, other virally or environmentally-induced
CC conditions, and certain congenital immune deficiencies. The peptides
CC can be employed to increase immune function which has been impaired by
CC use of immunosupressive drugs, such as certain chemotherapeutic drugs.
XX
SQ Sequence 766 AA;

Query Match 7.8%; Score 271; DB 15; Length 766;
Best Local Similarity 20.4%; Pred. No. 4.4e-14;
Matches 153; Conservative 113; Mismatches 285; Indels 200; Gaps 34;

```
Qy      15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
           |: |   :|||       : | | : |||::  ||         | :   :  |
Db     101 sindysispdgqfilleynyvkqwrhsytasydiydlnkrqliteeripnntqwv----- 155

Qy      66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
             ||| |  ||:: : :     :|:      : ::|     || |
Db     156 ----twspvghklayvwnnd-----iyvkiepnlpsyritwtgkediiyngitdwvyeee 206

Qy     110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
               |   |||:|  :           :| |    |   |    :|
Db     207 vfsaysalwwspngtflaya-------qfndtevplieysfysdeslq----------- 247

Qy     165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
             | : | |   |:|               | :|  |: |       ||  |
Db     248 ---ypktvrvpypkag--------------avnptvkffv-----vntdsls------- 277

Qy     224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
            |: :    |:|        :|      ||       :   ||   :|  ||
Db     278 -svtnatsiqit------apasmligdhylc------dvtwatqerislqwlrriqnysv 324

Qy     273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
```

```
Db     325 mdicdydessgrwnclvarqhiemstt----gwvgrfrpsephftldgnsfykiisneeg 380

Qy     326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366

Db     381 yrhicyfqidkkdwtfitkgtwevigiealtsdylyyiysneykgmpggrnlykiqlsdyt 440

Qy     367 ----------PDRPSELYSIPLGQEEK--QL-------------TGANDKFVR--EHTIS 399

Db     441 kvtclscelnper-cqyysvsfskeakyyqlrcsgpglplytlhssvndkglrvlednsa 499

Qy     400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449

Db     500 ldkmlqnvqmpskkldfiilnetkfwyqmilpphfdkskkyplplldvyagpcsqkadtvf 559

Qy     450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506

Db     560 rlnwatylasteniivasfdgrgsgyqqdkimhainrrlgtfevedqieaarqfskmg-f 618

Qy     507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565

Db     619 vdnkriaiwgwsyggyvtsmvlgsgsgvfkcgiavapvsrweyydsv------yterymg 672

Qy     566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623

Db     673 lptpednldhyrnstvmsraenfkqveyllihgtaddnvhfqqsaqiskalvdvgvdfqa 732

Qy     624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653

Db     733 mwytdedhgiasstahqhiythmshfikqcf 763
```

```
RESULT  14
AAW89596
ID   AAW89596 standard; Protein; 931 AA.
XX
AC   AAW89596;
XX
DT   17-MAR-1999  (first entry)
XX
DE   Saccharomyces cerevisiae dipeptidyl aminopeptidase I.
XX
KW   Prolyl dipeptidyl aminopeptidase; protein hydrolysate; dough;
KW   flavour enhancer; palatability; mouthfeel; aroma; crust colour;
KW   baking; animal feed additive; hydrolysis.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO9851803-A1.
XX
PD   19-NOV-1998.
XX
PF   12-MAY-1998;   98WO-US09629.
XX
PR   20-OCT-1997;   97US-0062892.
PR   16-MAY-1997;   97US-0857884.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
XX
PI   Blinkovsky A,  Brown K,  Byun T,  Klotz A,  Rey MW;
XX
DR   WPI; 1999-045232/04.
XX
PT   New dipeptidyl aminopeptidase from Aspergillus oryzae - used to
PT   produce protein hydrolysates enriched in particular amino acids,
PT   useful as flavour enhancers, e.g. in doughs
XX
PS   Example 7; Page 59-61; 77pp; English.
XX
CC   The present sequence represents dipeptidyl aminopeptidase (DPAP) from
CC   Saccharomyces cerevisiae. DPAP acts synergistically with aminopeptidase
CC   (AP) to hydrolyse polypeptides, producing protein hydrolysate (PH),
CC   useful in foods as flavour enhancer, e.g. in baked goods, enriched in:
CC   (a) Ala, Arg, Asp, Gly and/or Val, or (b), if the substrate has been
CC   deamidated, in Glu (free and/or peptide bound), in which case products
CC   are useful as animal feed additives. DPAP can also be used in flavour-
CC   improving compositions (optionally containing AP) and in dough pre-mixes,
CC   also for deactivating enzymes and for converting precursors to mature
CC   proteins. DPAP increases the level of hydrolysis of proteins and thus
CC   of flavour development, and a mixture with AP may hydrolyse tripeptides
CC   that are resistant to either enzyme used alone. PH have improved
CC   solubility, emulsifying and foaming properties, and products containing
CC   them have better flavour, palatability, mouthfeel, aroma and crust
CC   colour.
XX
SQ   Sequence   931 AA;
```

Query Match 7.7%; Score 269; DB 20; Length 931;
Best Local Similarity 20.1%; Pred. No. 8.9e-14;
Matches 155; Conservative 108; Mismatches 298; Indels 212; Gaps 30;

```
Qy       9 DITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDS--YTSNI-----------WIYETKTG 54

Db     220 detfevnlggnrflyegveftvstvqinykldklifgtnlesefrhsskgfywikdlntg 279

Qy      55 GSVPWTHGEKRSTD----------PRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKL 104

Db     280 niepilppeksddnyelglsklsyahfspaynyiyfvyen-----nlflqqvnsgvakkv 334

Qy     105 TDIPYGVSKPL--------------------WSPDGESILVTISLGEGESIDDREKTEQ 143

Db     335 te---dgskdifnakpdwiyeeevlasdqaiwwapddsk--avfarfndtsvddirlnry 389

Qy     144 DSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKEL-TSHKADHGDPAFSPDGKW 202

Db     390 tnmneaylsdtkikypkpgfqnpqf-dlflvnlqngiiysintgggkd----silyngkw 444

Qy     203 L---VFSANLTETDDASKPHDVYIMSLESG--------------------DLKQVTP-- 236

Db     445 ispdtfrfei--tdrnskildvkvydipssqmltvrntnsnlfngwiektkdilsippkp 502

Qy     237 --HRGSFGSSSFSPDGRYLALL----------------GNEKEYKNATLSKAWLYDIEQG 278

Db     503 elkrmdygyidihadsrgfshlfyyptvfakepiqltkgnwevtgngivg----yeyetd 558

Qy     279 RLTCLTEMLDV---HLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIE 335

Db     559 tifftaneigvmsqhlysisltdsttqntfqslqnpsdkydfy---------------- 601

Qy     336 GLVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRP------------SELYSIPLGQEE 382

Db     602 -------------dfelsssaryaiskklgpdtpikvagpltrvlnvaeihddsilqlt 647

Qy     383 KQLTGANDKF---VREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGG 439

Db     648 k-----dekfkekiknydlpitsyktmvlddgveinyieikpanlnpkkkypilvniygg 702

Qy     440 PHMMYGHTYF-------HEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDD 492

Db     703 p----gsqtfttksslafeqavvsgldvivlqieprgtggkgwsfrswareklgyweprd 758

Qy     493 VMQAVDEAIKRD-PHIDPKRLGVTGGSYGGFMTNWIVGQTN--RFKAAVTQRSISNWISF 549

Db     759 itevtkkfiqrnsqhideskiaiwgwsyggftslktveldngdtfkyamavapvtnwtly 818

Qy     550 HGV----------SDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598

Db     819 dsvyterymnqpsenhegyfevstiqnfksfeslkrlf---------------ivhgtf 862

Qy     599 DDRCPIEQAEQLFIALKKMG-KETKLVRFPNASHNLSRTGHPRQRI--KRLNY 648

Db     863 ddnvhiqntfrlvdqlnllgltnydmhifpdsdhsiry--hnaqrivfqklyy 913
```

RESULT 15

```
AAB18507
ID   AAB18507 standard; Protein; 732 AA.
XX
AC   AAB18507;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Amino acid sequence of prolyl-tripeptidyl peptidase PTP-A.
XX
KW   Prolyl tripeptidyl-peptidase; amidolytic activity; periodontal disease;
KW   gingivitis; periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200052147-A2.
XX
PD   08-SEP-2000.
XX
PF   03-MAR-2000; 2000WO-US05551.
XX
PR   05-MAR-1999;   99US-0123148.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
PA   (TRAV/) TRAVIS J.
PA   (POTE/) POTEMPA J.
PA   (BANB/) BANBULA A.
XX
PI   Travis J,  Potempa J,  Banbula A;
XX
DR   WPI; 2000-594181/56.
DR   N-PSDB; AAA75238.
XX
PT   Prolyl tripeptidyl-peptidase, active analog, fragment or variant useful
PT   for identifying its inhibitor which is useful for protecting an animal
PT   from a periodontal disease such as gingivitis and periodontitis -
XX
PS   Claim 6; Fig 7; 58pp; English.
XX
CC   The present sequence represents a prolyl tripeptidyl-peptidases from
CC   Porphyromonas gingivalis, which is designated PTP-A. The prolyl
CC   tripeptidyl-peptidase has an amidolytic activity, and cleaves a
CC   peptide bond in a target polypeptide having at least 4 amino acids.
CC   This bond is between a proline and an amino acid attached to the
CC   alpha-carboxyl group end of the proline. The polypeptide is useful
CC   for identifying inhibitors. These inhibitors are then useful for
CC   reducing the growth of bacterium or for protecting an animal from a
CC   periodontal disease such as gingivitis and periodontitis caused by
CC   Porphyromonas gingivalis.
XX
SQ   Sequence   732 AA;
```

```
  Query Match             7.5%;  Score 262.5;  DB 21;  Length 732;
  Best Local Similarity   22.2%;  Pred. No. 2.1e-13;
  Matches  159;  Conservative   85;  Mismatches  261;  Indels  211;  Gaps   31;

Qy      73 PDG----RTLAFISDREGDAAQ-LYIMSTEGG------EARKLTDIPYGVSKPLWSPDGE 121
           |:|     | || | |  || : | :: |:||       |||:| :
Db      97 pegckfqttdafpsfrtldagrglvvlftqgglvgfdmlarkvtyl-------------- 142

Qy     122 SILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRD----------GKGLTRG---- 166
                   |  |:|    : ||    ::| |:           |:|::|
Db     143 ------------fdtneetasldfspvg-drvayvrnhnlyiarggklgegmsraiavt 188

Qy     167 -------AYAQLVLVSVKSGEMKELTSHKADHG---DPAFSPDGKWLVF---------- 205
                 | | |              |: : |     :|| |  | |
Db     189 idgtetlvygqav-------------hqrefgiekgtfwspkgsclafyrmdqsmvkpt 234

Qy     206 -SANLTETDDASKP----------HDVYI----------MSLESGDLKQVTPHRGSFGSS 244
              :   :  |||          | | :          : |::|:     |       :
Db     235 pivdyhpleaeskplyypmagtpshhvtvgiyhlatgktvylqtge-----pkekfltnl 289

Qy     245 SFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRL--TCLTEMLDVHLADALIGDSLIG 302
           |:|||   | :    :         |   || | ||   |   |  | |  : |   : :
Db     290 swspdenilyvaevnraqnec---kvnaydaetgrfvrtlfve-tdkhyveplhpltflp 345

Qy     303 GAEQRPIWTKDSQGF---YV--------------------IGTDQGSTGIYYISIEGLV 338
           |:  : ||     |:   |:                     |  |   | :|: | |
Db     346 gsnnqfiwqsrrdgwnhlylydttgrlirqvtkgewevtnfagfdpkgtrlyfestea-- 403

Qy     339 YPIRLEKEY-----------------INSFSLSPDEQHFIASVTKPDRPSELYSIPLGQ 380
             ||: :                  ::   ||||    |      |  | ::    :|:
Db     404 --splerhfycidikggktkdltpesgmhrtqlspdgsaiidifqsptvprkvtvtnigk 461

Qy     381 EEKQL-------TGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLI 433
               |       ||      :|  ||       | :    :   |  |   :    ||:|
Db     462 gshtlleaknpdtgyampeirtgti-------maadgqtplyykltmplhfdpakkypvi 514

Qy     434 LNIHGGPHMMY----------GHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRG 483
           : ::|||||            |   :    :| |||||  :: |||   |  |   :
Db     515 vyvyggphaqlvtktwrssvggwdiy-----maqkgyavftvdsrgsanrgaafeqvihr 569

Qy     484 DYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFM-TNWIVGQTNRFKAAVTQRS 542
             |  :  | |  ||  :|    :|  |:|| | |||||| || ::   : ||  |
Db     570 rlgqtemadqmcgvd-flksqswvdadrigvhgwsyggfmttnlmlthgdvfkvgvaggp 628

Qy     543 ISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRC 602
           : :|  :  :    ||  |   |:    |   |     || | :::  |:::||  |
Db     629 vidwnryeimygeryf--dapqenpegydaanl-----lkragdlkgrlmlihgaidp-v 680

Qy     603 PIEQAEQLFI-ALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
           : |   ||: |  |          :|:  ||:        |:     |: :|  ||
Db     681 vvwqhsllfldacvkartypdyyvypshehnvmgp----drvhlyetitryftdhl 732
```

Search completed: September 25, 2001, 17:37:08
Job time: 135 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2001, 17:35:13 ; Search time 25.92 Seconds
(without alignments)
1930.813 Million cell updates/sec

Title: US-09-462-845-2
Perfect score: 3489
Sequence: 1 MKKLITADDITAIVSVTDPQ.....HPRQRIKRLNYISSWFDQHL 657

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3489	100.0	657	2	E70025	probable acylamino
2	1104.5	31.7	642	2	C71137	hypothetical prote
3	1096.5	31.4	631	2	H75007	probable acylamino
4	995	28.5	655	2	E75551	probable acyl-pept
5	962.5	27.6	659	2	F72568	probable acylamino
6	903	25.9	674	2	B84381	acylaminoacyl-pept
7	697	20.0	572	2	F72455	probable acylamino
8	667.5	19.1	632	2	E75057	peptidase PAB1418
9	660	18.9	622	2	F71174	hypothetical prote
10	473	13.6	709	2	B82580	alanyl dipeptidyl
11	410	11.8	591	2	H72474	probable acylamino
12	405.5	11.6	721	2	T09631	probable acylamino
13	396.5	11.4	732	1	S07624	acylaminoacyl-pept
14	391	11.2	732	1	JU0132	acylaminoacyl-pept
15	384	11.0	732	1	JC4655	acylaminoacyl-pept
16	357.5	10.2	741	2	JC5142	X-Pro dipeptidyl-p
17	355.5	10.2	598	2	F84199	hypothetical prote
18	350.5	10.0	745	2	T33751	hypothetical prote
19	346	9.9	637	2	S75772	hypothetical prote
20	334.5	9.6	606	2	T35378	probable peptide h
21	328.5	9.4	795	2	F82858	dipeptidyl-peptida
22	319.5	9.2	569	2	S74053	probable acylamino
23	319.5	9.2	761	2	S44807	F44B9.1 protein -
24	303	8.7	629	2	T15945	hypothetical prote
25	287.5	8.2	582	2	D72636	probable acylamino
26	278	8.0	711	2	S66261	X-Pro dipeptidyl-p
27	272	7.8	766	1	CDHU26	dipeptidyl-peptida
28	269	7.7	931	2	A49737	dipeptidyl aminope
29	243	7.0	760	1	S23752	dipeptidyl-peptida
30	240	6.9	829	2	T19514	hypothetical prote
31	239.5	6.9	792	1	A39914	dipeptidyl-peptida
32	238.5	6.8	450	2	H82151	tolB protein VC183
33	236.5	6.8	608	2	F83397	probable peptidase
34	226.5	6.5	931	2	T32919	hypothetical prote
35	224	6.4	803	2	I68600	dipeptidyl aminope
36	224	6.4	865	2	I54331	dipeptidyl aminope
37	223.5	6.4	714	2	E82507	hypothetical prote
38	223	6.4	779	2	T25173	hypothetical prote
39	223	6.4	799	2	T25174	hypothetical prote
40	221.5	6.3	655	2	T35975	probable acyl-pept
41	220	6.3	141	2	PD0003	acylaminoacyl-pept
42	219	6.3	759	2	I38593	fibroblast activat
43	218.5	6.3	818	1	A30107	dipeptidyl aminope
44	215	6.2	743	2	T37700	probable dipeptidy
45	215	6.2	955	2	F84914	hypothetical prote

ALIGNMENTS

RESULT 1
E70025
probable acylaminoacyl-peptidase (EC 3.4.19.1) yuxL - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: E70025; D25364
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: E70025
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-657 <KUN>
A;Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15213.1; PID:e11843
A;Experimental source: strain 168
R;Parsot, C.
EMBO J. 5, 3013-3019, 1986
A;Title: Evolution of biosynthetic pathways: a common ancestor for threonine synthase
A;Reference number: A91055; MUID:87080286
A;Accession: D25364
A;Molecule type: DNA
A;Residues: 'MPSVPQRQRRAPFFAGDKGVLLETNTCKGGDEYE',3-17,'S',19-24 <PAR>
A;Cross-references: GB:X04603; NID:g40210; PIDN:CAA28272.1; PID:g40213
A;Note: hypothetical fragment that is probably a mistranslation of a portion of the y
C;Genetics:
A;Gene: yuxL
C;Keywords: hydrolase; omega peptidase
F;518,599,631/Active site: Ser, Asp, His #status predicted

Query Match 100.0%; Score 3489; DB 2; Length 657;
Best Local Similarity 100.0%; Pred. No. 8.1e-226;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MKKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWT 60

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
```

```
Db      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120

Qy     121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180

Db     121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180

Qy     181 MKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGS 240

Db     181 MKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGS 240

Qy     241 FGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSL 300

Db     241 FGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSL 300

Qy     301 IGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHF 360

Db     301 IGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHF 360

Qy     361 IASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMR 420

Db     361 IASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMR 420

Qy     421 PAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNA 480

Db     421 PAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNA 480

Qy     481 VRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQ 540

Db     481 VRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQ 540

Qy     541 RSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDD 600

Db     541 RSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDD 600

Qy     601 RCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657

Db     601 RCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
```

RESULT 2
C71137
hypothetical protein PH0863 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: C71137
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: C71137
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-642 <KAW>
A;Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29957.1; PID:g3257274
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0863

Query Match 31.7%; Score 1104.5; DB 2; Length 642;
Best Local Similarity 39.9%; Pred. No. 3.7e-66;
Matches 255; Conservative 126; Mismatches 217; Indels 41; Gaps 21;

```
Qy      27 RAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRTLAFIS--DR 84

Db      35 KAVFQVTEISLKDDDYFSKLYLYDGKR--VKPFTSGNKDS-NPRFSPNGKLIAFTSKRDK 91

Qy      85 EGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQD 144

Db      92 EGKESELYVIPTDGGEARLLAKFKYGIKNLRFTEDGKSIAVVTPI-------DVEKKGND 144

Qy     145 SYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLV 204

Db     145 DVH--IIREIPFWFNGVGWIYGKRNVVYLVDVESGKKKRLTPKNLNVDQIRFH-NGR-LY 200

Qy     205 FSANLTETDDASKP--HDVYIMSLESGDLKQVTPHRGSFGSSSFSP--DGRYLALLGNEK 260

Db     201 FTA---QEDRERKPLISDLYV--LENRKVRKLTP--GKWRILDFLPLDDGSFV-LKANTL 252

Qy     261 EYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVI 320

Db     253 ERGIPTNAHIYHYDPKTGELKKLTKDLDRNAYNSLNSD--VRGSQRAELVYKEGWIYYV- 309

Qy     321 GTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIA--SVTKPDRPSELYSIPL 378

Db     310 ATDGPRANLFRVNLDGKIERVIGGDRSVESFDIG-DYIAFTAQDAVT----PTELYIYRD 364

Qy     379 GQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHG 438

Db     365 G-KEKKVTDFN-KWIKGYTLSKPEHFKVKASDGVEIDAWVMKPVNFRKGKKYPAILEIHG 422

Qy     439 GPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVD 498

Db     423 GPKTAYGYAFMHEFHVLTSKGFVVIFSNPRGSDGYGEEFAD-IRGHYGERDYQDLMEVVD 481

Qy     499 EAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYF 558

Db     482 EALRRFDFIDGERLGVTGGSYGGFMTNWIVGHTNRFKAAVTQRSISNWISFFGTTDIGYY 541

Qy     559 FTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMG 618

Db     542 FAPDQIGKDPWSNLEGYWEKSPLKYAPNVETPLLIIHSTEDYRCWLPEALQLFISLKYLG 601

Qy     619 KETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657

Db     602 KRVELAIFPGENHDLSRSGKPKHRVKRLELIAGWMEKWL 640
```

RESULT 3
H75007
probable acylaminoacyl-peptidase (EC 3.4.19.1) PAB1300 [similarity] - Pyrococcus abys
N;Alternate names: acyl-peptide hydrolase (APH); acylamino-acid-releasing enzyme
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 31-Mar-2000
C;Accession: H75007
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: H75007
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-631 <KAW>
A;Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1300
C;Keywords: hydrolase; omega peptidase

Query Match 31.4%; Score 1096.5; DB 2; Length 631;
Best Local Similarity 38.2%; Pred. No. 1.2e-65;
Matches 258; Conservative 119; Mismatches 233; Indels 65; Gaps 20;

```
Qy       1 MKKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSV-PW 59

Db       1 MMKKISEKDIEKFKLVGN--LDASGKKVVFQVTEISIKENDYFSSIYLYD---GRKVRRF 55

Qy      60 THGEKRSTDPRWSPDGRTLAFISDR--EGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWS 117

Db      56 TFGRK-DANPRFSPDGRLIAFTSKRSKEGRESELYLMPVDGGEARLLAKFKYGIKDLRFS 114

Qy     118 PDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRD----------GKGLT 164
```

```
Db    115 EDGKEIAV-------------------ITPVEVKKVANEKEDVHVIDEIPFWFNGIGWI 154

Qy    165 RGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKP--HDVY 222
Db    155 YGRRNALYLVDVESGKKRKVSPRNLDVLSVRFH-NGKLYL----LAQEDREKKPMISDVY 209

Qy    223 IMSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTC 282
Db    210 V--LEGKRVKKLTP--GKWRILDFVPLSDGLILKANTLERGFATTTSIYLY--RNGELRR 263

Qy    283 LTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIR 342
Db    264 LFKE-DLSAYNSLNSD--VRGAQRAELIFKDG-WIYFVATDGPRANLFRVNLDGKVERVI 319

Qy    343 LEKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPE 402
Db    320 RGDRSVESFSVG---DYIAFTAQDATTPTELYALRDG-KEKRITDFN-AWIRDYKLSKPE 374

Qy    403 EIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAV 462
Db    375 HFRVTASDGKEIDAWIMKPVDFKPGRKYPAVLEIHGGPKTAYGYAFMHEFQVLVAKGFVV 434

Qy    463 VYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGF 522
Db    435 IFSNPRGSDGYGEEFAD-IRGHYGERDYQDLMEVVDEALKRFDFIDEERLGVTGGSYGGF 493

Qy    523 MTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLK 582
Db    494 MTNWIVGHTNRFKAAVTQRSISNWISFFGTTDIGYYFAPDQIGGDPWSNLEGYWEKSPLK 553

Qy    583 YAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQR 642
Db    554 YAPNVETPLLIIHSTEDYRCWLPEALQFYIALKYLGKTVELAIFPGENHDLSRSGRPKHR 613

Qy    643 IKRLNYISSWFDQHL 657
Db    614 VKRLELIVGWFERWL 628
```

RESULT 4
E75551
probable acyl-peptide hydrolase - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: E75551
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: E75551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <WHI>
A;Cross-references: GB:AE001879; GB:AE000513; NID:g6457832; PIDN:AAF09754.1; PID:g645783
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0165
A;Map position: 1

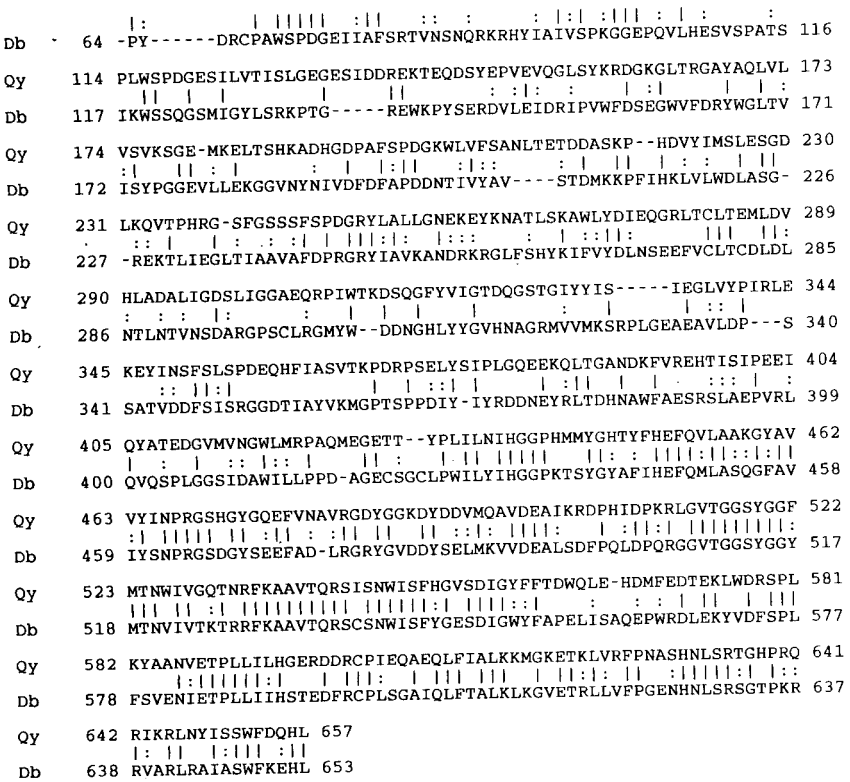
Query Match 28.5%; Score 995; DB 2; Length 655;
Best Local Similarity 36.0%; Pred. No. 8.1e-59;
Matches 250; Conservative 102; Mismatches 248; Indels 94; Gaps 19;

```
Qy      8 DDITAIVSVTDPQYAPDGTRAAYVKSQVNQE---------KDSYTSNIWIYETKTGGSV 57
Db     12 DSLLALAFPSDPQVSPDGKQVAFVLAQISEEDPAKPDKDFARPRYRSGLWLSE--GGAAR 69

Qy     58 PWTHGEKRSTD--PRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPL 115
Db     70 PLTHAETGRGDSAPRWSPDGQNLAFVRSAGEVKAALMLLPLKGGEARRVTHFKNGVSGPQ 129

Qy    116 WSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKG----LTRGAY--- 168
Db    130 WSPDGRFIAFT-------------TTADTED---------KRDERGEARVLTRPVYRAN 166

Qy    169 ---------AQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPH 219
Db    167 GADWLPERPAALWLYDVEADKLREWYAPEIGIGALSWWPDSRGVLIVQSEDEWQASQWRQ 226

Qy    220 DVYIMSLESGDLKQVTPHRGSFGSSSFS----PDGRYLALLGNEKEYKNATLSKAWLYDI 275
Db    227 DVYDLPLPTADAPAAPQKLLDWNSAAHGLAPHPDGQRFALIGRPAGKGNT--EHAHLYLI 284

Qy    276 EQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIE 335
Db    285 ENGQHRRLDTGHDHPVGDAVGGDCHVGAFPEGPRWL-DGDTLLFSSTVRGSVGLFTAHIG 343

Qy    336 GLVYPIRLEKEY-------INSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGA 388
Db    344 GGV------KAYDHDPQGVISAFTAN---EHGVALI----RESATRFPEVELNGQRVTDL 390

Qy    389 NDKF---VREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYG 445
Db    391 HARFPFPVRE-----PQRVTFETELG-EGEGWVLLP---EGEQKVPALLNIHGGPHTDYG 441

Qy    446 HTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDP 505
Db    442 HGFTHEFQLMAARGYGVCYSNPRGSVGYGQAWVDAIYGRWGTVDADDLLNFFDRCLEAVP 501

Qy    506 HIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
Db    502 RLDAAKTAVMGGSYGGFMTNWITGHTTRFQAAITDRCISNLISFGGTSDIGLRFWDDELG 561

Qy    566 HDMFE--DTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
Db    562 LDFSRRADALKLWDLSPLQYVENVKTPTLIVHSVLDHRCPVEQAEQWYAALHKHQVPVRF 621

Qy    624 VRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
Db    622 VRFPEENHELSRSGRPDRRLTRLNEYFAWLERWL 655
```

RESULT 5
F72568
probable acylamino-acid-releasing enzyme APE1832 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: F72568
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J. DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339
A;Accession: F72568
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-659 <KAW>
A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80835.1; PID:d1044621; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1832

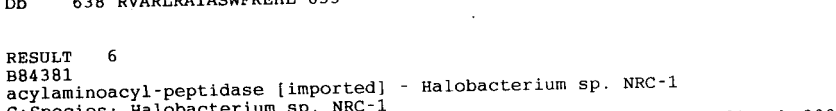
Query Match 27.6%; Score 962.5; DB 2; Length 659;
Best Local Similarity 34.8%; Pred. No. 1.2e-56;
Matches 235; Conservative 120; Mismatches 274; Indels 47; Gaps 18;

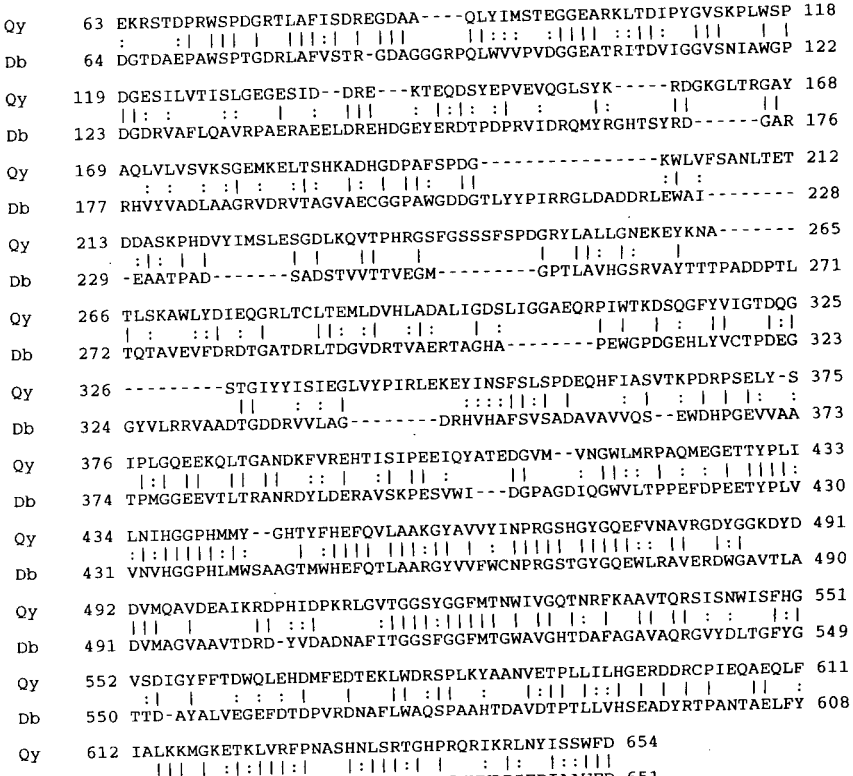
```
Qy      2 KKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWI-----YETKTGGS 56
Db      5 KKPVTPEDIIRLTFVSNPSVSPEGDKVAYLATKADEKENTYRSGIWLAEEDSYRPLTGG- 63

Qy     57 VPWTHGEKRSTDPRWSPDGRTLAF---ISDREGDAAQLYIMSTEGGEARKLTDIPYGVSK 113
```



RESULT 6
B84381
acylaminoacyl-peptidase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84381
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483
A;Accession: B84381
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-674 <STO>
A;Cross-references: GB:AE004437; NID:g10581715; PIDN:AAG20414.1; GSPDB:GN00138
C;Genetics:
A;Gene: yuxL

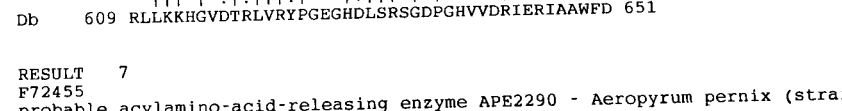
Query Match 25.9%; Score 903; DB 2; Length 674;
Best Local Similarity 31.7%; Pred. No. 1.2e-52;
Matches 223; Conservative 121; Mismatches 251; Indels 108; Gaps 22;





RESULT 7
F72455
probable acylamino-acid-releasing enzyme APE2290 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: F72455
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339
A;Accession: F72455
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-572 <KAW>
A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81302.1; PID:d1045088; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE2290

Query Match 20.0%; Score 697; DB 2; Length 572;
Best Local Similarity 33.5%; Pred. No. 5.9e-39;
Matches 199; Conservative 99; Mismatches 248; Indels 48; Gaps 20;



```
Db      2 PSWSPDASKLAFASRRGSGKGEKGSGLYIVG-RGGEPRRVAWFTHGVSELGWA-SATKVY 59

Qy    125 VTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKEL 184

Db     60 VVAPVAD-------EKTRDKDGDYVATDRLPLWFDSSGLVAGLADQIHLVDAESGRTSRL 112

Qy    185 TSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGSFGSS 244

Db    113 TSE--DLGVLTAEPCGN-SIYYVTLRRWAD---PLDTVVKSVPISGGEPETVLEGLTVSQ 166

Qy    245 SFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLT-EMLDVHLADALIGDSLIGG 303

Db    167 LRCINGK-LFMLAHKREIGLASHSKLYLLE-DRGKASCLTCGILDRNIW------SIAGG 218

Qy    304 AEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIAS 363

Db    219 LGGEPV---------IVYADRGRSVI--AAVEGRVEDLVRGDMIVVQADSNGEEVVYIAS 267

Qy    364 VTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQ 423

Db    268 --SPTEPPEVYRFKGGDVER-VSSINRWVVEEFRLYKPRRVEVEAE-GEVVEGWYIEP-- 321

Qy    424 MEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRG 483

Db    322 -EGEGRKPLILFIHGGPKGMYGYGFYGEMQLFASEGFTVAYANPRGSDGYEEEFAD-IRG 379

Qy    484 DYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSI 543

Db    380 RYGDTDYKQLMKFLD-AVVEGSDVDEGKMAVTGISYGGYMTNVIVTKTSRFRAAVSENGI 438

Qy    544 SNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCP 603

Db    439 ADWIADFWAADIGYWFDPDQIGGTPLDNLEEYVKRSPAFHAKNVNTPVMLIHSMEDYRCF 498

Qy    604 IEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657

Db    499 IDQALAMHTALKVNGKESFLVVFTKGSHGHSILASPRHRRKRLELKLKWIKEKL 552
```

RESULT 8
E75057
peptidase PAB1418 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: E75057
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: E75057
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-632 <KAW>
A;Cross-references: GB:AJ248287; GB:AL096836; NID:g5458657; PIDN:CAB50354.1; PID:e151625
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1418

Query Match 19.1%; Score 667.5; DB 2; Length 632;
Best Local Similarity 28.4%; Pred. No. 6.5e-37;
Matches 194; Conservative 132; Mismatches 263; Indels 93; Gaps 27;

```
Qy      1 MKKLITADDI----------TAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY 49

Db      1 MMQLIFGDTVMTSIEWNEKTFTKFAYLSDPRIR--GSTIAYVLTKANLDNNKYESTI-VL 57

Qy     50 ETKTGGSVPWTHGEKR----STDPRWSPDGRTLAFIS-DREGDAAQLYIMSTEGGEARKL 104

Db     58 ENLEDGS-------KRFIEDASMPRISPDGKKIAFMRFNEEKKVAQIWVADMKTLSAKKV 110

Qy    105 TDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLT 164
```

```
Db    111 LEAK-NIRSLEWNDDSRRLLV---------IGFKRRDDEDFIFEDDVPAWF---DNMGFF 157

Qy    165 RGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKW----LVFSANLTET-DDASKPH 219

Db    158 DGEKTTFWIVDTEAEEVIE------QFEKPRFS-SGVWHGDSIVVNVPYRDTIPRYFKYW 210

Qy    220 DVYIMSLESGDLKQVTPHRGSFGSSSF---SPDGRYLALLGN-EKEYKNATLSKAWLYDI 275

Db    211 NIYLW--KDGEEEKL------FEKVSFHAIDSDGKNILLYGKPEKKYMSEH-DKLYIYD- 260

Qy    276 EQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIE 335

Db    261 --GEVKGIMDSIDREAGQAKI--------------KDGKVYFTL-YEEGSVNLYL--WD 300

Qy    336 GLVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVRE 395

Db    301 GDIKEIAKGKHWIMGFDVDEKVVYLKETAT---RLRELYI--WDGEERQLTDYNGLIFSK 355

Qy    396 HTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVL 455

Db    356 LKTFEPKHFKFKSID-LEIDGWYIKPELKEGEKA-PVIVFVHGGPKGMYGYYFKYEMQLM 413

Qy    456 AAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVT 515

Db    414 ASKGYYIVYVNPRGSNGYSEDFALRVLERTGLEDFQDILNGIEEFFKLEPQADRERVGIT 473

Qy    516 GGSYGGFMTNWIVGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKL 575

Db    474 GISYGGFMTNWALTQSDLFKAGISENGISYWLTSYAFSDIGLWFDKEVIGENPLEN-ENF 532

Qy    576 WDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSR 635

Db    533 RKLSPLFYAKNVKAPLLLIHSLEDYRCPLDQSLMFYHVLKDLGKEVYIAIFKRGAHGHSI 592

Qy    636 TGHPRQRIKRLNYISSWFDQHL 657

Db    593 RGSPRHRMKRYKLFMEFFERKL 614
```

RESULT 9
F71174
hypothetical protein PH0594 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: F71174
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137
A;Accession: F71174
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-622 <KAW>
A;Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29683.1; PID:g3257000
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;Gene: PH0594

Query Match 18.9%; Score 660; DB 2; Length 622;
Best Local Similarity 28.2%; Pred. No. 2e-36;
Matches 189; Conservative 130; Mismatches 251; Indels 100; Gaps 26;

```
Qy     11 TAIVSVTDPQYAPDGTR---AAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRST 67

Db     12 TKFAYLSDPR-----TRKNLVAYVLTKANLESNKYENTI-VIENLEDGSRKFI---EDAS 62

Qy     68 DPRWSPDGRTLAFIS-DREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVT 126

Db     63 MPRISPDGKKIAFMRFNEEKKTAQIWVADLKTLSAKKVLEAK-NIRSIEWNQDSRRLL-- 119
```

```
Qy    127 ISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTS 186
Db    120 -------AVGFKRREDEDFIFEDDVPAWF---DNMGFFDGEKTTFWVIDTEGEEVIE--- 166

Qy    187 HKADHGDPAFSPDGKW----LVFSANLTETDDASKPH----------DVYIMSLESGDL 231
Db    167 ---QFEKPRFS-SGIWHGDSIVVSV----------PHRDVIPRYFKYWDIYLW--KDGEE 210

Qy    232 KQVTPHRGSFGSSSF---SPDGRYLALLGN-EKEYKNATLSKAWLYDIEQGRLTCLTEML 287
Db    211 EKL------FEKVSFYAIDSDGERILLYGKPEKKYVSEH-DKIYIYD---GEVKGILDDI 260

Qy    288 DVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEY 347
Db    261 DREVAQAKIRNGKV----------------YFTLFEEGSVNLYL--WDGEVREIAKGKHW 302

Qy    348 INSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYA 407
Db    303 IMGFDA---DERLIYLKETATRPAELYL--WDGEERQLTDYNGLIFKKLKTFEPRHFRFK 357

Qy    408 TEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINP 467
Db    358 SID-LELDGWYIKPEIKEGEKA-PVIVFVHGGPKGMYGYYFKYEMQLMASKGYYIVYVNP 415

Qy    468 RGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWI 527
Db    416 RGSNGYSEDFALRVLERTGLEDFQDILNGIEEFLRLEPQADRERIGITGISYGGYMTNWA 475

Qy    528 VGQTNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANV 587
Db    476 LTQSDLFKAGISENGISYWLTSYAFSDIGLWF-DKEVIGDNPLENENYRKLSPLFYAKNV 534

Qy    588 ETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLN 647
Db    535 KAPLLLIHSLEDYRCPLDQSLMFYHVLKDLGKEVYIAIFKKGAHGHSIRGSPRHRMKRYK 594

Qy    648 YISSWFDQHL 657
Db    595 LFMEFFERKL 604
```

RESULT 10
B82580
alanyl dipeptidyl peptidase XF2260 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: B82580
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82580
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-709 <SIM>
A;Cross-references: GB:AE004038; GB:AE003849; NID:g9107408; PIDN:AAF85059.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigi
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:

A;Gene: XF2260

Query Match 13.6%; Score 473; DB 2; Length 709;
Best Local Similarity 22.7%; Pred. No. 8.2e-24;
Matches 163; Conservative 138; Mismatches 288; Indels 130; Gaps 23;

```
Qy      9 DITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGS--------VPWT 60
Db     49 DMVALDRVSSPELSPDGAVLVFAKRQMDAKYIKASTSVWVQRLQAGTSAAPVRLTPLGWD 108

Qy     61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
Db    109 -----VSAPVFSRDGKAVYFLSAKSG-SHQLYVLPVSGGTSRQLTNLAVDIDSYKLSPQG 162

Qy    121 ESILVTISL----GEGESIDDREKTEQDS-------YEPVEVQGLSYKRDGKGLTRGAYA 169
Db    163 DRIVFSAGVFQVCGSDLSCTKRKLDEKKNAKASGVVFEQLFVRHWDTWNDGRRNTL-FIA 221

Qy    170 QLVLVSVK------------SGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDA-S 216
Db    222 SLPAVGAKPVSVVSAMSAMLDGDVPSKPFGGADHF--VWSPDGHSVVASIRVAGRQEPWS 279

Qy    217 KPHDVYIMSLESGD----LKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWL 272
Db    280 TNFDLYRFDVSGHDAPVNLTAANPAWDA--TPMFSADGKMLYYRAMRRPGFEADRFGLME 337

Qy    273 YDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTK---------DSQGFYVIGTD 323
Db    338 MEVQSGKV----------------------REIAPHWDRSADEIALSADGKALYVNADD 374

Qy    324 QGSTGIYYISI----------EGLVYPIRLE--KEYINSFSLSPDEQHFIASVTKPDRPS 371
Db    375 HGEHPLFKVDIASGKVEKWVGEGSVHAPVLAGGKLAFARNSLKSADQIFVTDAVARG--- 431

Qy    372 ELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYP 431
Db    432 -----PL----QAITSATGEVLQQVRLGDFEQFSFKGWNDETVYGYVVKPYDYQPGKKYP 482

Qy    432 LILNIHGGPHMMYGHTYFHEF--QVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKD 489
Db    483 VAFLIHGGPQGSFGNSWGYRWNPQTYAGQGYAVVMIDFHGSTGYGQAFTDAISQHWGDRP 542

Qy    490 YDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQRSISNW--- 546
Db    543 LEDLQKGWAAAQQQYPFLNGDKACALGASYGGYMVYWIAGHWNQ----------PWKCL 591

Qy    547 ISFHGVSD---IGY-----FFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
Db    592 VDHDGVFDNRMMGYATEELWFSEWENGGTPWENPAGYEQFNPVLHVDKWRVPMLVIHGQK 651

Qy    599 DDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
Db    652 DFRIPVEQGLAAFGALQRKGIESKLLYFHDENHWVLK---PQNSIQWHDTVNAWLKKYI 707
```

RESULT 11
H72474
probable acylamino-acid-releasing enzyme APE2441 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: H72474
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339
A;Accession: H72474
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-591 <KAW>
A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81456.1; PID:d1045242; PID:g
A;Experimental source: strain K1

C;Genetics:
A;Gene: APE2441

Query Match 11.8%; Score 410; DB 2; Length 591;
Best Local Similarity 24.9%; Pred. No. 1e-19;
Matches 167; Conservative 99; Mismatches 244; Indels 162; Gaps 26;

```
Qy      12 AIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTDPRW 71
Db       8 AVKSSHTPRIDP-STGCIYYLS------DSASSQPVIWKSCESRNDVWLPWERRVGSLEI 60

Qy      72 SPDGRTLAFISDREGDAA-QLYIMSTE------GGEARKLTDI-PYGVSKPLWSPDGESI 123
Db      61 APDG-LVAFSTDKDGDERWSIYIAGEDLGVRLVAGEDGSINMIGP-------WSPDGRLL 112

Qy     124 LVTISLGEGESID----DREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSG 179
Db     113 AFTSNKRNGVDFDLYVFDREK----------GSVSIVVEGEGIIAAS---------- 149

Qy     180 EMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRG 239
Db     150 --------------------SWIDGSRLLAVKRNSNLDSDILAVNVYKGEAKVLTRHRG 188

Qy     240 SFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDS 299
Db     189 EELNTSPRPIGGGKALFISNMD--------------------SEFTGIALIDLQTGDR 226

Qy     300 LIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEG--LVYPIRLEKEYI--------- 348
Db     227 RLVFREQ---W--DVEALEV----SGKTVVYSLNVDGESRVYTTKIDERGSLLQPRPIEG 277

Qy     349 ----NSFSLSPDEQHFIA--SVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPE 402
Db     278 LPRGSLLSLDLNEKLGLAVFSLSTPKHGIEIYIAGLGRSASRLTVSPKAWLREDEFVEPE 337

Qy     403 EIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEF-QVLAAKGYA 461
Db     338 HFSYESFDGIKVRALLYKP-DKPLYTPPPAVVYLHGGPESQ-ERVRFNVFPQALAAIGIA 395

Qy     462 VVYINPRGSHGYGQEFV---------NAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRL 512
Db     396 TIAPNYRGSTGYGRRFVHLDDVEKRMDAVR---------DVYYAVKAAVEAG-LVDGSRL 445

Qy     513 GVTGGSYGGFMTNWIVG-QTNRFKAAVTQRSISNWISF----------HGVSDIGYFFTD 561
Db     446 CVMGGSYGGYLTLMSLAIYPDLWKCGVEIVGIVNLVTFIRNTSPYRRRYRIAEYG----- 500

Qy     562 WQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKET 621
Db     501 -----DPDVHGEIMLKLSPITYVENMKAPLMVIHGAKDPRVPVSEAEQLVEALSSRGVRV 555

Qy     622 KLVRFPNASHNL 633
Db     556 RYVRLEDEGHGI 567
```

RESULT 12
T09631
probable acylaminoacyl-peptidase (EC 3.4.19.1) - Aspergillus fumigatus
C;Species: Aspergillus fumigatus
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C;Accession: T09631
R;Beauvais, A.; Monod, M.; Svab, J.; Kobayashi, H.; Diaquin, M.; Hovanessian, A.G.; Latg
J. Biol. Chem. 272, 6238-6244, 1997
A;Title: Biochemical and antigenic characterization of a new dipeptidyl-peptidase isolat
A;Reference number: Z16787; MUID:97197790
A;Accession: T09631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-721 <BEA>
A;Cross-references: EMBL:L48074; NID:g2340045; PID:g2340046
C;Genetics:
A;Introns: 22/2; 39/3; 117/2; 453/3; 528/1; 605/3; 676/2
C;Keywords: hydrolase; omega peptidase

Query Match 11.6%; Score 405.5; DB 2; Length 721;
Best Local Similarity 22.3%; Pred. No. 2.8e-19;
Matches 161; Conservative 107; Mismatches 298; Indels 155; Gaps 26;

```
Qy       6 TADDITAIVSVTDPQYA-----PDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWT 60
Db      15 TALALTPEQLITAPRRSEAIPDPSGKVAVFSTSQYSFETHKRTSWWSLLDLKTGQTKVLT 74

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGD---AAQLYI---------------------- 93
Db      75 N-DSSVSEIVWLSDD-SILYVNSTNADIPGGVELWVTQASSFAKGYKAASLPASFSGLKA 132

Qy      94 MSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQG 153
Db     133 AKTKSGDIR---FVAYGQSYPNGTAYNEE-LATAPLSSARIYDSIYVRHWDYWLSTTFNA 188

Qy     154 LSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTS--HKADHGDPAF--------SPDGKWL 203
Db     189 VFSGTLKKGHGKNGYS-------LDGELKNLVSPVKNAESPYPPFGGASDYDLSPDGKWV 241

Qy     204 VFSANLTETDDAS--------KPHDVYIMSLE-SGDLKQVTPH--RGSFGSSSFSPDGRY 252
Db     242 AFKSKAPELPKANFTTSYIYLVPHDASETARPINGPDSPGTPKGIKGDSSSPVFSPNGDK 301

Qy     253 LALLG-NEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRP--- 308
Db     302 LAYFQMRDETYES---DRALLYVYSLGSKKTI----------PSVAGDWDRSPDSV 344

Qy     309 IWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEY----------INSFSLSPDEQ 358
Db     345 KWTPDGKTLIVGSEDLGRTRLFSL-------PANAKDDYKPKNFTDGGSVSAYYFLPDSS 397

Qy     359 HFIAS----------VTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYAT 408
Db     398 LLVTGSALWTNWNVYTAKPEKGVIKKIASANEIDPELKGLGPSDISEF---------YFQ 448

Qy     409 EDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEF--QVLAAKGYAVVYIN 466
Db     449 GNFTDIHAWVIYPENFDKSKKYPLIFFIHGGPQGNWADGWSTRWNPKAWADQGYVVVAPN 508

Qy     467 PRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNW 526
Db     509 PTGSTGFGQALTTAIQNNWGGAPYDDLVKCWEYVHENLDYVDTDHGVAAGASYGGFMINW 568

Qy     527 IVGQT--NRFKAAVTQRSISNWISFHGVSDIGYFFTD--------WQLEHDM------FE 570
Db     569 IQGSPLGRKFKAL----------VSHDGTFVADAKVSTEELWFMQREFNGTFWDAR 614

Qy     571 DTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNAS 630
Db     615 DNYRRWDPSAPERILQFATPMLVIHSDKDYRLPVAEGLSLFNVLQERGVPSRFLNFPDEN 674

Qy     631 H 631
Db     675 H 675
```

RESULT 13
S07624
acylaminoacyl-peptidase (EC 3.4.19.1) - rat
N;Alternate names: acyl-peptide hydrolase; acylamino-acid-releasing enzyme
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 18-Jun-1999
C;Accession: A33706; S07624
R;Kobayashi, K.; Lin, L.W.; Yeadon, J.E.; Klickstein, L.B.; Smith, J.A.
J. Biol. Chem. 264, 8892-8899, 1989
A;Title: Cloning and sequence analysis of a rat liver cDNA encoding acyl-peptide hydr
A;Reference number: A33706; MUID:89255359
A;Accession: A33706

```
A;Molecule type: mRNA
A;Residues: 1-732 <KOB>
A;Cross-references: GB:J04733; NID:g202931; PIDN:AAA88506.1; PID:g202932
A;Note: parts of this sequence, including the amino end of the mature protein, were dete
R;Lin, L.W.; Lee, F.J.S.; Smith, J.A.
Nucleic Acids Res. 17, 4397-4400, 1989
A;Title: Structural organization of the rat acyl-peptide hydrolase gene.
A;Reference number: S07624; MUID:89296508
A;Accession: S07624
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-732 <LIN>
A;Cross-references: EMBL:X14915
C;Genetics:
A;Introns: 4/3; 49/1; 91/2; 122/3; 148/1; 202/3; 248/3; 279/2; 293/1; 333/3; 354/1; 386/
C;Superfamily: acylaminoacyl-peptidase
C;Keywords: blocked amino end; homotetramer; hydrolase; omega peptidase
F;1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental
F;118,291,443/Modified site: lysine derivative (Lys) #status experimental
F;587,675,707/Active site: Ser, Asp, His #status predicted

  Query Match             11.4%;  Score 396.5;  DB 1;  Length 732;
  Best Local Similarity   21.7%;  Pred. No. 1.1e-18;
  Matches  148;  Conservative  126;  Mismatches  271;  Indels  137;  Gaps   21;

Qy      22 APDGTRAAYVK---SQVNQEKDSYTSNIW------------------IYETKTGGSVPW 59
           :| ||  | ::     |: |:  :   :|                    :||    | : |
Db      97 SPSGTMKAVLRKAGGTVSGEEKQFL-EVWEKNRKLKSFNLSALEKHGPVYEDDCFGCLSW 155

Qy      60 THGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPD 119
           :| |           | ::::::                                    |
Db     156 SHSETH-----------LLYVAEKK------------------------------RPK 172

Qy     120 GESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSG 179
            ||   | :|    | |:  : ::       | : |:  |: :   :   | :: : ||
Db     173 AESFFQTKALDISASDDEMARPKKPDQAIKGDQFVFYEDWGETMVSKSIPVLCVLDIDSG 232

Qy     180 EMKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDL 231
           : |         |   ::|    :||          |       ::   :| : |  |
Db     233 NISVLEGVPENVSPGQAFWAPGDTGVVFVGWWHEPFRLGIRYCTNRRSALYYVDLSGGKC 292

Qy     232 KQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHL 291
           : ::    :  |  |||   :  |       :   |:  |||          | :| :: | :
Db     293 ELLSDGSLAICSPRLSPDQCRIVYLQYPCLAPHHQCSQLCLYD----WYTKVTSVV-VDI 347

Qy     292 ADALIGDSLIG-GAEQRPI--WTKDS------------QGFYVIGTDQGSTGIYYISIEG 336
           :|:|  |      |:  |: ||            |  : : |  ||  |  ::  |
Db     348 VPRQLGESFSGIYCSLLPLGCWSADSQRVVFDSAQRSRQDLFAVDTQTGS--ITSLTAAG 405

Qy     337 LVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSEL---YSIPLGQEE----------K 383
              :|      |: |:   :|  : |  |  |   :  | |:|:           :
Db     406 SAGSWKL---------LTIDKDLMVAQFSTPSLPPSLKVGFLPPPGKEQSVSWVSLEEAE 456

Qy     384 QLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMM 443
           : | :      |     | :|||  |   :   |::|:    :|  |:::  |||||
Db     457 PIPGIHWGVRVLHPPPDQENVQYADLDFEAI---LLQPSNPPDKTQVPMVVMPHGGPHSS 513

Qy     444 YGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKR 503
           :   :     :|   |:||: :| ||| |:||: : :: |: | :|  ||  ||::  ::
Db     514 FVTAWMLFPAMLCKMGFAVLLVNYRGSTGFGQDSILSLPGNVGHQDVKDVQFAVEQVLQE 573

Qy     504 DPHIDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDW 562
           : | | :|: : |||:|||::  ::||    : | : :  :  |  | :||     ||
Db     574 E-HFDARRVALMGGSHGGFLSCHLIGQYPETYSACIARNPVINIASMMGSTDI----PDW 628

Qy     563 QLEHDMFEDT----------EKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFI 612
           :    |  :           |:: |:||:||   |:||:|:: |: | | | :|  : :
Db     629 CMVETGFPYSNSCLPDLNVWEEMLDKSPIKYIPQVKTPVLLMLGQEDRRVPFKQGMEYYR 688

Qy     613 ALKKMGKETKLVRFPNASHNLS 634
           |||      :|: :| ::| ||
Db     689 ALKARNVPVRLLLYPKSNHALS 710
```

```
RESULT  14
JU0132
acylaminoacyl-peptidase (EC 3.4.19.1) [validated] - pig
N;Alternate names: acyl-peptide hydrolase; acylamino-acid-releasing enzyme
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 21-Jul-2000
C;Accession: JU0132
R;Mitta, M.; Asada, K.; Uchimura, Y.; Kimizuka, F.; Kato, I.; Sakiyama, F.; Tsunasawa
J. Biochem. 106, 548-551, 1989
A;Title: The primary structure of porcine liver acylamino acid-releasing enzyme deduc
A;Reference number: JU0132; MUID:90110044
A;Accession: JU0132
A;Molecule type: mRNA
A;Residues: 1-732 <MIT>
C;Genetics:
A;Introns: 4/3; 49/1; 91/2; 122/3; 148/1; 202/3; 248/3; 279/2; 293/1; 333/3; 354/1; 3
C;Function:
A;Description: EC 3.4.19.1 [validated; MUID:90110044]
C;Superfamily: acylaminoacyl-peptidase
C;Keywords: acetylated amino end; homotetramer; hydrolase; omega peptidase
F;1/Modified site: acetylated amino end (Met) #status experimental
F;587,675,707/Active site: Ser, Asp, His #status predicted

  Query Match             11.2%;  Score 391;  DB 1;  Length 732;
  Best Local Similarity   21.0%;  Pred. No. 2.7e-18;
  Matches  149;  Conservative  123;  Mismatches  287;  Indels  152;  Gaps   23;

Qy      22 APDGTRAAYVKSQ--VNQEKDSYTSNIW------------------IYETKTGGSVPWT 60
           :| ||  | ::        ::     :|                    :||    | : |:
Db      97 SPSGTMKAVLRKAGGTGTAEEKQFLEVWEKNRKLKSFNLSALEKHGPVYEDDCFGCLSWS 156

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
           | |                 | ::::::                              |
Db     157 HSETH-----------LLYVAEKK------------------------------RPKA 173

Qy     121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180
            ||   | :|      | |:  :|::        | | |:  |: :   :   | :: ::||
Db     174 ESFFQTKALDVTGSDDEMARTKKPDQAIKGDQFLFYEDWGENMVSKSTPVLCVLDIESGN 233

Qy     181 MKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDLK 232
           : |         |   ::|    :||          |       ::   :| : |  | :
Db     234 ISVLEGVPESVSPGQAFWAPGDTGVVFVGWWHEPFRLGIRFCTNRRSALYYVDLTGGKCE 293

Qy     233 QVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLA 292
           ::    :  |  |||   : |         :   : |||      | :| :: | :
Db     294 LLSDESVAVTSPRLSPDQCRIVYLRFPSLVPHQQCGQLCLYD----WYTRVTSVV-VDIV 348

Qy     293 DALIGDSLIG-GAEQRPI--WTKDS------------QGFYVIGTDQGSTGIYYISIEGL 337
           :|:  |      |:  |: ||            |  : : |  ||  |  ::  |
Db     349 PRQLGEDFSGIYCSLLPLGCWSADSQRVVFDSPQRSRQDLFAVDTQMGSVTSLTAGGSGG 408

Qy     338 VYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSEL---YSIPLGQEEKQLTGANDKFVR 394
           : :          |: |    :   : |  |  |   :  | |:|:     |
Db     409 SWKL----------LTIDRDLMVVQFSTPSVPPSLKVGFLPPAGKEQ-----AVSWVSL 452

Qy     395 EHTISIP--------------EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGG 439
           |   |                | :|||   |:     |::|:    :|  |:::  |||
Db     453 EEAEPFPDISWSIRVLQPPPQQEHVQYA---GLDFEAILLQPSNSPEKTQVPMVVMPHGG 509

Qy     440 PHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDE 499
           ||  :   :     :|   |:||: :| ||| |:||: : :: |: | :|  ||  ||::
Db     510 PHSSFVTAWMLFPAMLCKMGFAVLLVNYRGSTGFGQDSILSLPGNVGHQDVKDVQFAVEQ 569

Qy     500 AIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYF 558
           :: : | |  |: : |||:|||::  ::||    : | :  :  |  | :||
Db     570 VLQEE-HFDAGRVALMGGSHGGFLSCHLIGQYPETYSACVVRNPVINIASMMGSTDI--- 625

Qy     559 FTDWQLEHDMFEDTE------KLW----DRSPLKYAANVETPLLILHGERDDRCPIEQAE 608
           || :    | :         :|    |:||:|||  |:||||:: |: | | | :|.
Db     626 -PDWCMVEAGFSYSSDCLPDLSVWAAMLDKSPIKYAPQVKTPLLLMLGQEDRRVPFKQGM 684
```

```
Qy    609 QLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISS--WFDQHL 657
Db    685 EYYRVLKARNVPVRLLLYPKSTHALSEV-----EVESDSFMNAVLWLCTHL 730


RESULT  15
JC4655
acylaminoacyl-peptidase (EC 3.4.19.1) - human
N;Alternate names: acidpeptide hydrolase; acylamino acid releasing enzyme; lung DNF15S2
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JC4655; A42257; A30145; A53799
R;Mitta, M.; Ohnogi, H.; Mizutani, S.; Sakiyama, F.; Kato, I.; Tsunasawa, S.
DNA Res. 3, 31-35, 1996
A;Title: The nucleotide sequence of human acylamino acid-releasing enzyme.
A;Reference number: JC4655; MUID:96281126
A;Accession: JC4655
A;Molecule type: mRNA
A;Residues: 1-732 <MIT>
A;Cross-references: DDBJ:D38441; NID:g556513; PIDN:BAA07476.1; PID:g556514
A;Experimental source: liver
R;Scaloni, A.; Jones, W.M.; Barra, D.; Pospischil, M.; Sassa, S.; Popowicz, A.; Manning,
J. Biol. Chem. 267, 3811-3818, 1992
A;Title: Acylpeptide hydrolase: inhibitors and some active site residues of the human en
A;Reference number: A42257; MUID:92156118
A;Accession: A42257
A;Molecule type: protein
A;Residues: 4-18;45-53;90-95;119-126;130-136,'A',138-149;173-193,'T',195-199,'A',201-219
71,'Q',473-475,'V',477-497;519-527;539-574,'M',576-591,'L',593;656-657,'K',659-664;677-6
A;Experimental source: lung
R;Naylor, S.L.; Marshall, A.; Hensel, C.; Martinez, P.F.; Holley, B.; Sakaguchi, A.Y.
Genomics 4, 355-361, 1989
A;Title: The DNF15S2 locus at 3p21 is transcribed in normal lung and small cell lung can
A;Reference number: A30145; MUID:89233127
A;Accession: A30145
A;Molecule type: mRNA
A;Residues: 102-112,'D',114-136,'A',138-167,'K',169-196,'QTKPSRGISLCFMKTGRNMVSKSIPVSACWM
ALQIRPKCPWWSCPT',509,'A',511-581,'X',583-604,'RACLRTRDHNASMLGSTDILTGAWW','RL',617-618,63
A;Cross-references: GB:J03068; NID:g181629; PIDN:AAA35769.1; PID:g181630
A;Experimental source: lung
R;Scaloni, A.; Barra, D.; Jones, W.M.; Manning, J.M.
J. Biol. Chem. 269, 15076-15084, 1994
A;Title: Human acylpeptide hydrolase. Studies on its thiol groups and mechanism of actio
A;Reference number: A53799; MUID:94253066
A;Accession: A53799
A;Molecule type: protein
A;Residues: 19-33;63-65 <SC2>
C;Genetics:
A;Gene: GDB:APEH; AARE; D3S48E; DNF15S2
A;Cross-references: GDB:127917; OMIM:102645
A;Map position: 3p21.3-3p21.2
C;Function:
A;Description: hydrolyzes amino-terminal acetylated peptides into peptides having free a
C;Superfamily: acylaminoacyl-peptidase
C;Keywords: acetylated amino end; blocked amino end; homotetramer; hydrolase; liver; ome
F;1/Modified site: acetylated amino end (Met) #status predicted
F;587/Active site: Ser #status experimental
F;675,707/Active site: Asp, His #status predicted


  Query Match             11.0%;  Score 384;  DB 1;  Length 732;
  Best Local Similarity   21.0%;  Pred. No. 7.8e-18;
  Matches  147;  Conservative  136;  Mismatches  287;  Indels  130;  Gaps   24;

Qy     22 APDGTRAAYVKS----------------QVNQEKDSYTSNIW-----IYETKTGGSVPWT 60
Db     97 SPSGSMKAVLRKAGGTGPGEEKQFLEVWEKNRKLKSFNLSVLEKHGPVYEDDCFGCLSWS 156

Qy     61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
Db    157 HSETH------------LLYVAERK------------------------------RPKA 173

Qy    121 ESILVTISLGEGESIDDREKTEQDSYEPVE-VQGLSYKRDGKGLTRGAYAQLVLVSVKSG 179
Db    174 ESFFQTKALDVSASDDEIARLKKPD-QPIKGDQFVFYEDWGENMVSKSIPVLCVLDVESG 232

Qy    180 EMKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDL 231
Db    233 NISVLEGVPENVSPGQAFWAPGDAGVVFVGWWHEPFRLGIRFCTNRRSALYYVDLIGGKC 292

Qy    232 KQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHL 291
Db    293 ELLSDDSLAVSSPRLSPDQCRIVYLQYPSLIPHHQCSQLCLYD----WYTKVTSVV-VDV 347

Qy    292 ADALIGDSLIG-GAEQRPI--WTKDSQGFYVIGTDQGSTGIYYISIE-GLVYPIRLEKEY 347
Db    348 VPRQLGENFSGIYCSLLPLGCWSADSQRVVFDSAQRSRQDLFAVDTQVGTVTSLTAGGSG 407

Qy    348 INSFSLSPDEQHFIASVTKPDRPSELY--SIPLGQEEKQLTGANDKFVREHTISIP---- 401
Db    408 GSWKLLTIDQDLMVAQFSTPSLPPTLKVGFLPSAGKEQSVLWVS----LEEAEPIPDIHW 463

Qy    402 -----------EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFH 450
Db    464 GIRVLQPPPEQENVQYA---GLDFEAILLQPGSPPDKTQVPMVVMPHGGPHSSFVTAWML 520

Qy    451 EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPK 510
Db    521 FPAMLCKMGFAVLLVNYRGSTGFGQDSILSLPGNVGHQDVKDVQFAVEQVLQEE-HFDAS 579

Qy    511 RLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDW------- 562
Db    580 HVALMGGSHGGFISCHLIGQYPETYRACVARNPVINIASMLGSTDI----PDWCVVEAGF 635

Qy    563 QLEHDMFEDTE---KLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGK 619
Db    636 PFSSDCLPDLSVWAEMLDKSPIRYIPQVKTPLLLMLGQEDRRVPFKQGMEYYRALKTRNV 695

Qy    620 ETKLVRFPNASHNLSRTGHPRQRIKRLNYISS--WFDQHL 657
Db    696 PVRLLLYPKSTHALSEV-----EVESDSFMNAVLWLRTHL 730


Search completed: September 25, 2001, 17:37:39
Job time: 146 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2001, 17:35:43 ; Search time 16.84 Seconds
(without alignments)
1336.452 Million cell updates/sec

Title: US-09-462-845-2
Perfect score: 3489
Sequence: 1 MKKLITADDITAIVSVTDPQ.....HPRQRIKRLNYISSWFDQHL 657

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3489	100.0	657	1	YUXL_BACSU	P39839 bacillus su
2	396.5	11.4	732	1	ACPH_RAT	P13676 rattus norv
3	395	11.3	732	1	ACPH_PIG	P19205 sus scrofa
4	384	11.0	732	1	ACPH_HUMAN	P13798 homo sapien
5	319.5	9.2	761	1	YL31_CAEEL	P34422 caenorhabdi
6	272	7.8	766	1	DPP4_HUMAN	P27487 homo sapien
7	269	7.7	931	1	DAP1_YEAST	P33894 saccharomyc
8	247	7.1	760	1	DPP4_MOUSE	P28843 mus musculu
9	242.5	7.0	767	1	DPP4_RAT	P14740 rattus norv
10	233	6.7	859	1	DPP6_RAT	P46101 rattus norv
11	224	6.4	865	1	DPP6_HUMAN	P42658 homo sapien
12	218.5	6.3	818	1	DAP2_YEAST	P18962 saccharomyc
13	211.5	6.1	863	1	DPP6_BOVIN	P42659 bos taurus
14	190	5.4	427	1	TOLB_HAEIN	P44677 haemophilus
15	179.5	5.1	710	1	PPCE_PIG	P23687 sus scrofa
16	177.5	5.1	710	1	PPCE_HUMAN	P48147 homo sapien
17	174.5	5.0	432	1	TOLB_PSEAE	P50601 pseudomonas
18	170	4.9	430	1	TOLB_ECOLI	P19935 escherichia
19	168.5	4.8	443	1	TOLB_RICPR	Q9zdm5 rickettsia
20	160.5	4.6	689	1	PPCE_AERHY	Q06903 aeromonas h
21	151	4.3	722	1	Y174_RICPR	P81171 rickettsia
22	148	4.2	1356	1	HET1_PODAN	Q00808 podospora a
23	142.5	4.1	726	1	Y4NA_RHISN	P55577 rhizobium s
24	142	4.1	291	1	DLHH_METEX	P71505 methylobact
25	138.5	4.0	705	1	PPCF_FLAME	P27195 flavobacter
26	137.5	3.9	1693	1	Y163_SYNY3	Q55563 synechocyst
27	136.5	3.9	431	1	TOLB_CHLPN	Q9z7c4 chlamydia p
28	136	3.9	416	1	YIDR_ECOLI	P31455 escherichia
29	135.5	3.9	705	1	PPCE_FLAME	P27028 flavobacter
30	131.5	3.8	614	1	YA3A_SCHPO	Q09715 schizosacch
31	128.5	3.7	348	1	YE91_SYNY3	P74598 synechocyst
32	126.5	3.6	431	1	TOLB_CHLTR	O84604 chlamydia t
33	124	3.6	281	1	YN07_MYCTU	Q50658 mycobacteri
34	124	3.6	1191	1	Y143_SYNY3	P74442 synechocyst
35	121.5	3.5	561	1	EST4_RAT	Q64573 rattus norv
36	121	3.5	5147	1	FAT_DROME	P33450 drosophila
37	120.5	3.5	1068	1	TRI_SULSO	P95871 sulfolobus
38	117.5	3.4	326	1	YE09_SYNY3	P73594 synechocyst
39	116	3.3	438	1	SHPI_STAHY	Q08002 staphylococ
40	115.5	3.3	246	1	DLHH_SYNY3	P73163 synechocyst
41	115.5	3.3	302	1	RE14_SCHPO	Q09150 schizosacch
42	115	3.3	1065	1	SED4_YEAST	P25365 saccharomyc
43	114.5	3.3	514	1	TUP1_CANAL	P56093 candida alb
44	114.5	3.3	583	1	T2F1_FLAOK	P14870 flavobacter
45	114	3.3	428	1	PSG3_HUMAN	Q16557 homo sapien

ALIGNMENTS

```
RESULT   1
YUXL_BACSU
ID   YUXL_BACSU     STANDARD;      PRT;   657 AA.
AC   P39839; O32120;
DT   01-FEB-1995 (Rel. 31, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROBABLE PEPTIDASE YUXL (EC 3.4.21.-).
GN   YUXL.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=87080286; PubMed=3098560;
RA   Parsot C.;
RT   "Evolution of biosynthetic pathways: a common ancestor for threonine
RT   synthase, threonine dehydratase and D-serine dehydratase.";
RL   EMBO J. 5:3013-3019(1986).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z99120; CAB15213.1; -.
DR   EMBL; X04603; CAA28272.1; ALT_FRAME.
DR   PIR; D25364; D25364.
DR   SubtiList; BG10463; yuxL.
KW   Hypothetical protein; Hydrolase; Serine protease.
FT   ACT_SITE    518    518       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    631    631       CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ   SEQUENCE   657 AA;  73713 MW;  5A46AA40D5F6EAF9 CRC64;
```

Query Match 100.0%; Score 3489; DB 1; Length 657;
Best Local Similarity 100.0%; Pred. No. 1.9e-218;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MKKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKKLITADDITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWT 60

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
```

```
Db     61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120

Qy    121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180

Db    121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180

Qy    181 MKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGS 240

Db    181 MKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGS 240

Qy    241 FGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSL 300

Db    241 FGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSL 300

Qy    301 IGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHF 360

Db    301 IGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHF 360

Qy    361 IASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMR 420

Db    361 IASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMR 420

Qy    421 PAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNA 480

Db    421 PAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNA 480

Qy    481 VRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQ 540

Db    481 VRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQTNRFKAAVTQ 540

Qy    541 RSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDD 600

Db    541 RSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDD 600

Qy    601 RCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657

Db    601 RCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
```

```
RESULT   2
ACPH_RAT
ID   ACPH_RAT       STANDARD;      PRT;   732 AA.
AC   P13676; P14320; P70479;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ACYLAMINO-ACID-RELEASING ENZYME (EC 3.4.19.1) (ACYL-PEPTIDE HYDROLASE)
DE   (APH) (ACYLAMINOACYL-PEPTIDASE).
GN   APEH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89255359; PubMed=2722805;
RA   Kobayashi K., Lin L.-W., Yeadon J.E., Klickstein L.B., Smith J.A.;
RT   "Cloning and sequence analysis of a rat liver cDNA encoding acyl-
RT   peptide hydrolase.";
RL   J. Biol. Chem. 264:8892-8899(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=89296508; PubMed=2578023;
RA   Lin L.-W., Lee F.J.S., Smith J.A.;
RT   "Structural organization of the rat acyl-peptide hydrolase gene.";
RL   Nucleic Acids Res. 17:4397-4400(1989).
CC   -!- FUNCTION: THIS ENZYME CATALYZES THE HYDROLYSIS OF THE AMINO-
CC       TERMINAL PEPTIDE BOND OF AN N-ACETYLATED PEPTIDE TO GENERATE
CC       AN N-ACETYLATED AA AND A PEPTIDE WITH A FREE AMINO-TERMINUS.
CC       IT PREFERENTIALLY CLEAVES OFF AC-ALA, AC-MET AND AC-SER.
CC   -!- CATALYTIC ACTIVITY: ACYLAMINOACYL-PEPTIDE + H(2)O = ACYLAMINO
CC       ACID + PEPTIDE.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9C; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04733; AAA88506.1; -.
DR   EMBL; X14915; CAA33040.1; ALT_SEQ.
DR   PIR; S07624; S07624.
DR   MEROPS; S09.004; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Acetylation.
FT   MOD_RES       1      1       ACETYLATION (POTENTIAL).
FT   MOD_RES     118    118       CHEMICAL NATURE NOT DETERMINED.
FT   MOD_RES     291    291       CHEMICAL NATURE NOT DETERMINED.
FT   MOD_RES     443    443       CHEMICAL NATURE NOT DETERMINED.
FT   ACT_SITE    587    587       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    675    675       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    707    707       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CONFLICT    625    628       IPDW -> M (IN REF. 2).
SQ   SEQUENCE   732 AA;  81383 MW;  43F234879E10235B CRC64;
```

```
  Query Match             11.4%;  Score 396.5;  DB 1;  Length 732;
  Best Local Similarity   21.7%;  Pred. No. 2.6e-18;
  Matches  148;  Conservative  126;  Mismatches  271;  Indels  137;  Gaps   21;

Qy      22 APDGTRAAYVK---SQVNQEKDSYTSNIW------------------IYETKTGGSVPW 59

Db      97 SPSGTMKAVLRKAGGTVSGEEKQFL-EVWEKNRKLKSFNLSALEKHGPVYEDDCFGCLSW 155

Qy      60 THGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPD 119

Db     156 SHSETH-----------LLYVAEKK-----------------------------RPK 172

Qy     120 GESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSG 179

Db     173 AESFFQTKALDISASDDEMARPKKPDQAIKGDQFVFYEDWGETMVSKSIPVLCVLDIDSG 232

Qy     180 EMKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDL 231

Db     233 NISVLEGVPENVSPGQAFWAPGDTGVVFVGWWHEPFRLGIRYCTNRRSALYYVDLSGGKC 292

Qy     232 KQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHL 291

Db     293 ELLSDGSLAICSPRLSPDQCRIVYLQYPCLAPHHQCSQLCLYD----WYTKVTSVV-VDI 347

Qy     292 ADALIGDSLIG-GAEQRPI--WTKDS-----------QGFYVIGTDQGSTGIYYISIEG 336

Db     348 VPRQLGESFSGIYCSLLPLGCWSADSQRVVFDSAQRSRQDLFAVDTQTGS--ITSLTAAG 405

Qy     337 LVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSEL---YSIPLGQEE----------K 383

Db     406 SAGSWKL---------LTIDKDLMVAQFSTPSLPPSLKVGFLPPPGKEQSVSWVSLEEAE 456

Qy     384 QLTGANDKFVREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMM 443

Db     457 PIPGIHWGVRVLHPPPDQENVQYADLDFEAI---LLQPSNPPDKTQVPMVVMPHGGPHSS 513

Qy     444 YGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKR 503

Db     514 FVTAWMLFPAMLCKMGFAVLLVNYRGSTGFGQDSILSLPGNVGHQDVKDVQFAVEQVLQE 573
```

```
Qy     504 DPHIDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDW 562
           : | | :|: : |||:|||::  ::||    : | : :  : |  |  | :||     ||
Db     574 E-HFDARRVALMGGSHGGFLSCHLIGQYPETYSACIARNPVINIASMMGSTDI----PDW 628

Qy     563 QLEHDMFEDT----------EKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFI 612
             :   | :             |:: |:||:||   |:||:|:: |: | | | :|  : :
Db     629 CMVETGFPYSNSCLPDLNVWEEMLDKSPIKYIPQVKTPVLLMLGQEDRRVPFKQGMEYYR 688

Qy     613 ALKKMGKETKLVRFPNASHNLS 634
           |||      :|: :| ::| ||
Db     689 ALKARNVPVRLLLYPKSNHALS 710


RESULT   3
ACPH_PIG
ID   ACPH_PIG       STANDARD;      PRT;   732 AA.
AC   P19205;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ACYLAMINO-ACID-RELEASING ENZYME (EC 3.4.19.1) (ACYL-PEPTIDE HYDROLASE)
DE   (APH) (ACYLAMINOACYL-PEPTIDASE).
GN   APEH.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90110044; PubMed=2691504;
RA   Mitta M., Asada K., Uchimura Y., Kimizuka F., Kato I., Sakiyama F.,
RA   Tsunasawa S.;
RT   "The primary structure of porcine liver acylamino acid-releasing
RT   enzyme deduced from cDNA sequences.";
RL   J. Biochem. 106:548-551(1989).
CC   -!- FUNCTION: THIS ENZYME CATALYZES THE HYDROLYSIS OF THE AMINO-
CC       TERMINAL PEPTIDE BOND OF AN N-ACETYLATED PEPTIDE TO GENERATE
CC       AN N-ACETYLATED AA AND A PEPTIDE WITH A FREE AMINO-TERMINUS.
CC       IT PREFERENTIALLY CLEAVES OFF AC-ALA, AC-MET AND AC-SER.
CC   -!- CATALYTIC ACTIVITY: ACYLAMINOACYL-PEPTIDE + H(2)O = ACYLAMINO
CC       ACID + PEPTIDE.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9C; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00524; BAA00411.1; -.
DR   PIR; JU0132; JU0132.
DR   MEROPS; S09.004; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Acetylation.
FT   MOD_RES      1      1       ACETYLATION.
FT   ACT_SITE    587    587      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    675    675      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    707    707      CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ   SEQUENCE   732 AA;  81243 MW;  017BD40E049A604C CRC64;


 Query Match           11.3%;  Score 395;  DB 1;  Length 732;
 Best Local Similarity  21.1%;  Pred. No. 3.2e-18;
 Matches  150;  Conservative  122;  Mismatches  287;  Indels  152;  Gaps   23;

Qy      22 APDGTRAAYVKSQ--VNQEKDSYTSNIW----------------IYETKTGGSVPWT 60
           :| ||  | ::       ::     :|                    :||    | : |:
Db      97 SPSGTMKAVLRKAGGTGTAEEKQFLEVWEKNRKLKSFNLSALEKHGPVYEDDCFGCLSWS 156

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
           | |              | :::|::                                |
Db     157 HSETH----------LLYVADKK----------------------------RPKA 173

Qy     121 ESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGE 180
           ||   | :|    | |:  :|::        | | |:  |: :   :  | :: ::||
Db     174 ESFFQTKALDVTGSDDEMARTKKPDQAIKGDQFLFYEDWGENMVSKSTPVLCVLDIESGN 233

Qy     181 MKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDLK 232
           : |          |   ::|    :||         |       ::   :| : |  |  :
Db     234 ISVLEGVPESVSPGQAFWAPGDTGVVFVGWWHEPFRLGIRFCTNRRSALYYVDLTGGKCE 293

Qy     233 QVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLA 292
           ::    :  |  |||   : |        :    :  |||      | :| :: | :
Db     294 LLSDESVAVTSPRLSPDQCRIVYLRFPSLVPHQQCGQLCLYD----WYTRVTSVV-VDIV 348

Qy     293 DALIGDSLIG-GAEQRPI--WTKDS-----------QGFYVIGTDQGSTGIYYISIEGL 337
             :|:   |      |:  |: ||             |  : : |  ||         |
Db     349 PRQLGEDFSGIYCSLLPLGCWSADSQRVVFDSPQRSRQDLFAVDTQMGSVTSLTAGGSGG 408

Qy     338 VYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSEL---YSIPLGQEEKQLTGANDKFVR 394
           : :           |: |    :  : | |  |   :  | |:|:     |
Db     409 SWKL----------LTIDRDLMVVQFSTPSVPPSLKVGFLPPAGKEQ-----AVSWVSL 452

Qy     395 EHTISIP--------------EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGG 439
           |     |                | :|||  |:     |::|:    :|  |:::  |||
Db     453 EEAEPFPDISWSIRVLQPPPQQEHVQYA---GLDFEAILLQPSNSPEKTQVPMVVMPHGG 509

Qy     440 PHMMYGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDE 499
           ||  :   :    :|   |:||: :| ||| |:||: : :: |: | :|  ||  ||::
Db     510 PHSSFVTAWMLFPAMLCKMGFAVLLVNYRGSTGFGQDSILSLPGNVGHQDVKDVQFAVEQ 569

Qy     500 AIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYF 558
           :: : | |  |: : |||:|||::  ::||    : | | :  : |  |  | :||
Db     570 VLQEE-HFDAGRVALMGGSHGGFLSCHLIGQYPETYSACVVRNPVINIASMMGSTDI--- 625

Qy     559 FTDWQLEHDMFEDTE------KLW----DRSPLKYAANVETPLLILHGERDDRCPIEQAE 608
             || :   |  :         :|    |:||:|||  |:||||:: |: | | | :|
Db     626 -PDWCMVEAGFSYSSDCLPDLSVWAAMLDKSPIKYAPQVKTPLLLMLGQEDRRVPFKQGM 684

Qy     609 QLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISS--WFDQHL 657
           : :  ||      :|: :| ::| ||        ::  :::::  |  ||
Db     685 EYYRVLKARNVPVRLLLYPKSTHALSEV-----EVESDSFMNAVLWLCTHL 730


RESULT   4
ACPH_HUMAN
ID   ACPH_HUMAN     STANDARD;      PRT;   732 AA.
AC   P13798;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ACYLAMINO-ACID-RELEASING ENZYME (EC 3.4.19.1) (ACYL-PEPTIDE HYDROLASE)
DE   (APH) (ACYLAMINOACYL-PEPTIDASE) (DNF15S2 PROTEIN).
GN   APEH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Mitta M., Ohnogi H., Mizutani S., Kato I., Sakiyama F., Tsunasawa S.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   PRELIMINARY SEQUENCE OF 102-732 FROM N.A.
RX   MEDLINE=89233127; PubMed=2565880;
RA   Naylor S.L., Marshall A., Hensel C., Martinez P.F., Holley B.,
```

```
RA   Sakaguchi A.Y.;
RT   "The DNF15S2 locus at 3p21 is transcribed in normal lung and small
RT   cell lung cancer.";
RL   Genomics 4:355-361(1989).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=91319413; PubMed=1861871;
RA   Erlandsson R., Boldog F., Persson B., Zabarovsky E.R., Allikmets R.L.,
RA   Sumegi J., Klein G., Joernvall H.;
RT   "The gene from the short arm of chromosome 3, at D3F15S2, frequently
RT   deleted in renal cell carcinoma, encodes acylpeptide hydrolase.";
RL   Oncogene 6:1293-1295(1991).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=91172778; PubMed=2006156;
RA   Jones W.M., Scaloni A., Bossa F., Popowicz A.M., Schneewind O.,
RA   Manning J.M.;
RT   "Genetic relationship between acylpeptide hydrolase and acylase, two
RT   hydrolytic enzymes with similar binding but different catalytic
RT   specificities.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:2194-2198(1991).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE=94016571; PubMed=8411161;
RA   Feese M., Scaloni A., Jones W.M., Mannig J.M., Remington S.J.;
RT   "Crystallization and preliminary X-ray studies of human erythrocyte
RT   acylpeptide hydrolase.";
RL   J. Mol. Biol. 233:546-549(1993).
CC   -!- FUNCTION: THIS ENZYME CATALYZES THE HYDROLYSIS OF THE AMINO-
CC       TERMINAL PEPTIDE BOND OF AN N-ACETYLATED PEPTIDE TO GENERATE
CC       AN N-ACETYLATED AA AND A PEPTIDE WITH A FREE AMINO-TERMINUS.
CC       IT PREFERENTIALLY CLEAVES OFF AC-ALA, AC-MET AND AC-SER.
CC   -!- CATALYTIC ACTIVITY: ACYLAMINOACYL-PEPTIDE + H(2)O = ACYLAMINO
CC       ACID + PEPTIDE.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9C; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY DUE TO
CC       FRAMESHIFT ERRORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D38441; BAA07476.1; -.
DR   EMBL; J03068; AAA35769.1; ALT_FRAME.
DR   MEROPS; S09.004; -.
DR   MIM; 102645; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Acetylation.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   ACT_SITE    587    587       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    675    675       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    707    707       CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ   SEQUENCE   732 AA;  81292 MW;  3842FB353CE78A40 CRC64;


  Query Match             11.0%;  Score 384;  DB 1;  Length 732;
  Best Local Similarity   21.0%;  Pred. No. 1.7e-17;
  Matches  147;  Conservative  136;  Mismatches  287;  Indels  130;  Gaps   24;

Qy      22 APDGTRAAYVKS----------------QVNQEKDSYTSNIW-----IYETKTGGSVPWT 60
           :| |:  | ::                 : |::  |:  ::      :||    | : |:
Db      97 SPSGSMKAVLRKAGGTGPGEEKQFLEVWEKNRKLKSFNLSVLEKHGPVYEDDCFGCLSWS 156

Qy      61 HGEKRSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
```

```
           | |              | ::::|:                                    |
Db     157 HSETH-----------LLYVAERK-----------------------------RPKA 173

Qy     121 ESILVTISLGEGESIDDREKTEQDSYEPVE-VQGLSYKRDGKGLTRGAYAQLVLVSVKSG 179
           ||   | :|    | |:  : ::   :|::  | : |:  |: :   :   | :: |:||
Db     174 ESFFQTKALDVSASDDEIARLKKPD-QPIKGDQFVFYEDWGENMVSKSIPVLCVLDVESG 232

Qy     180 EMKEL--TSHKADHGDPAFSPDGKWLVFSA------NLTETDDASKPHDVYIMSLESGDL 231
           :  |          |   ::|    :||         |      ::   :| : |  |
Db     233 NISVLEGVPENVSPGQAFWAPGDAGVVFVGWWHEPFRLGIRFCTNRRSALYYVDLIGGKC 292

Qy     232 KQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHL 291
           : ::    :  |   |||   : |          :   |:  |||      | :| :: | :
Db     293 ELLSDDSLAVSSPRLSPDQCRIVYLQYPSLIPHHQCSQLCLYD----WYTKVTSVV-VDV 347

Qy     292 ADALIGDSLIG-GAEQRPI--WTKDSQGFYVIGTDQGSTGIYYISIE-GLVYPIRLEKEY 347
           :|:: |      |:  |: |||        :    :: :  : | |  :
Db     348 VPRQLGENFSGIYCSLLPLGCWSADSQRVVFDSAQRSRQDLFAVDTQVGTVTSLTAGGSG 407

Qy     348 INSFSLSPDEQHFIASVTKPDRPSELY--SIPLGQEEKQLTGANDKFVREHTISIP---- 401
           :  |: |:  :|  : |  |  |    :|  :|: :   :     |   ||
Db     408 GSWKLLTIDQDLMVAQFSTPSLPPTLKVGFLPSAGKEQSVLWVS----LEEAEPIPDIHW 463

Qy     402 ----------EEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFH 450
                    | :|||   |:     |::|     :|  |::: |||||  :   :
Db     464 GIRVLQPPPEQENVQYA---GLDFEAILLQPGSPPDKTQVPMVVMPHGGPHSSFVTAWML 520

Qy     451 EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPK 510
           :|   |:|: :| ||| |:||: : :: |: | :|  ||  ||:: :: : | |
Db     521 FPAMLCKMGFAVLLVNYRGSTGFGQDSILSLPGNVGHQDVKDVQFAVEQVLQEE-HFDAS 579

Qy     511 RLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDW------- 562
           : : |||:|||::  ::||    ::| | :  : | |  | :||     ||
Db     580 HVAIMGGSHGGFISCHLIGQYPETYRACVARNPVINIASMLGSTDI----PDWCVVEAGF 635

Qy     563 QLEHDMFEDTE---KLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGK 619
           |   |     :: |:||::|   |:|||:: |: | | | :|  : : |||
Db     636 PFSSDCLPDLSVWAEMLDKSPIRYIPQVKTPLLLMLGQEDRRVPFKQGMEYYRALKTRNV 695

Qy     620 ETKLVRFPNASHNLSRTGHPRQRIKRLNYISS--WFDQHL 657
             :|: :| ::| ||        ::  ::::::  |   ||
Db     696 PVRLLLYPKSTHALSEV-----EVESDSFMNAVLWLRTHL 730


RESULT   5
YL31_CAEEL
ID   YL31_CAEEL     STANDARD;      PRT;   761 AA.
AC   P34422;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 86.0 KDA PROTEIN F44B9.1 IN CHROMOSOME III.
GN   F44B9.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
```

```
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9C; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY. STRONG, TO ACYLAMINO-ACID-RELEASING
CC       ENZYME.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L23648; AAA28026.1; -.
DR   PIR; S44807; S44807.
DR   WormPep; F44B9.1; CE00522.
DR   InterPro; IPR002471; -.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE_NEG.
KW   Hypothetical protein; Hydrolase.
FT   ACT_SITE    537    537       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    625    625       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    657    657       CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ   SEQUENCE   761 AA;  85963 MW;  7766F0AAF2650364 CRC64;


  Query Match             9.2%;  Score 319.5;  DB 1;  Length 761;
  Best Local Similarity   20.4%;  Pred. No. 2.6e-13;
  Matches  143;  Conservative  129;  Mismatches  247;  Indels  183;  Gaps   28;

Qy      68 DPRW-----SPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-DIPYGVSKPLWSPDGE 121
           ||::     ||| : : :::  |     ::   :    :|::| :  : :   :|:   :
Db      50 DPKYSSVSLSPDAKQVGYVAPDENGIRNVFTRCSSCSYSRQVTFETEHPILNYVWTAIPD 109

Qy     122 SILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYA----QLVLVSVK 177
           || |                ||::     :   ||   | ::  | |    | |::| |
Db     110 VILFT---------------QDNHGDENTR--IYK---KNISATAIAADKTQRVVISEK 148

Qy     178 SGEMKELTSHKADHGDPAFSPDGKWLVFSANL-TET------DDASKPHDVYIMSLESGD 230
                               |  | :: | || :||      |:    |::|  : :: :
Db     149 -------------------PMVKAMILSNNLISETVLIGMNDENPALHNIYAFNCQTDE 188

Qy     231 LKQVTPHRGSFG-----------SSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQG 278
           || |  :|  |             :|   |||  :      |   |   |:
Db     189 LKLVLQNR-RFSIFFFDNDLNVRLASEEGPDGEMIYYRPRSNEGARTTEQNTWV------ 241

Qy     279 RLTCLTEMLDVHLADA----------------LIGD-SLIGGAEQRPIWTKDSQGFYVI 320
           | | :  |                      ::|| | :|     |   :| |   ::
Db     242 ------EYLRIQHDDKAITMPITFDKSNNFMYWIMGDGSDLGNLVVFPF--EDPQQKEIL 293

Qy     321 GTDQGS----------------TGIYY---------ISIEGLVYPIRLEKE-YINSFSL 353
           | | :                | :|:           :| | | : ::    :|  |:
Db     294 YTAQRAQIGNVLIHPTDKTLLAVTEVYHKPELFVANETFMEDLQYLVNMKPSGSMNIVSM 353

Qy     354 SPDEQHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATEDGVM 413
           | |   :: : :  | | ::|      :: :|  :    ::::|::      :   | :
Db     354 SIDMSTWLVTYSSSDEPYDIYLYRRWNKKAELFMSTRPELKKYTLNKQIGFDFRARDEMT 413

Qy     414 VNGWLMRPAQM---------EGETTY-----------PLILNIHGGPHMM--YGHTYFHE 451
           :  :|  | |           :|:  |            :|: :|||     || :  :
Db     414 IQAYLSLPPQAPLLKSSQVPDGDRPYANLGMIPAVPQKMIVLVHGGPKARDHYGFSPMNA 473

Qy     452 FQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKR 511
           :  |  :||:|: :| ||| |:|:   ||  |::| | : |:: ||: |: :    :
Db     474 W--LTNRGYSVLQVNFRGSTGFGKRLTNAGNGEWGRKMHFDILDAVEFAVSKG-IANRSE 530

Qy     512 LGVTGGSYGGFMTNW-------------IVGQTNRFKAAVTQRSISNWISFHG--VSDI 555
           : | ||||||: |                ||| :|    :: |     |: |    :  :
Db     531 VAVMGGSYGGYETLVALTFTPQTFACGVDIVGPSNLI--SLVQAIPPYWLGFRKDLIKMV 588

Qy     556 GYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALK 615
           |   :|        |  : |  |||| :|  |  |::|: |  | |    :::|   ||:
Db     589 GADISD-------EEGRQSLQSRSPLFFADRVTKPIMIIQGANDPRVKQAESDQFVAALE 641

Qy     616 KMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
           |        : :|:  | : :   |:  :::  :| ::  | |
Db     642 KKHIPVTYLLYPDEGHGVRK---PQNSMEQHGHIETFLQQCL 680


RESULT   6
DPP4_HUMAN
ID   DPP4_HUMAN     STANDARD;      PRT;   766 AA.
AC   P27487;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DIPEPTIDYL PEPTIDASE IV (EC 3.4.14.5) (DPP IV) (T-CELL ACTIVATION
DE   ANTIGEN CD26) (TP103) (ADENOSINE DEAMINASE COMPLEXING PROTEIN-2)
DE   (ADABP).
GN   DPP4 OR ADCP2 OR CD26.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92329551; PubMed=1352704;
RA   Misumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
RT   "Molecular cloning and sequence analysis of human dipeptidyl
RT   peptidase IV, a serine proteinase on the cell surface.";
RL   Biochim. Biophys. Acta 1131:333-336(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=95012454; PubMed=7927537;
RA   Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
RT   "Genomic organization, exact localization, and tissue expression of
RT   the human CD26 (dipeptidyl peptidase IV) gene.";
RL   Immunogenetics 40:331-338(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood;
RX   MEDLINE=92325476; PubMed=1352530;
RA   Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
RA   Dahlberg H.N., Schlossman S.F., Morimoto C.;
RT   "Cloning and functional expression of the T cell activation antigen
RT   CD26.";
RL   J. Immunol. 149:481-486(1992).
RN   [4]
RP   ERRATUM.
RX   MEDLINE=93171637; PubMed=8094732;
RA   Tanaka T.;
RL   J. Immunol. 150:2090-2090(1993).
RN   [5]
RP   SEQUENCE OF 1-551 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=92165847; PubMed=1347043;
RA   Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
RA   Trotot P., Barbat A.;
RT   "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
RT   colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
RT   human coding sequence and changes of dipeptidyl peptidase IV mRNA
RT   levels during cell differentiation.";
RL   J. Biol. Chem. 267:4824-4833(1992).
RN   [6]
RP   SEQUENCE OF 545-766 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91024044; PubMed=1977364;
RA   Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
RT   "Isolation of a cDNA probe for the human intestinal
RT   dipeptidylpeptidase IV and assignment of the gene locus DPP4 to
RT   chromosome 2.";
RL   Ann. Hum. Genet. 54:191-197(1990).
RN   [7]
RP   SEQUENCE OF 1-31 FROM N.A.
```

```
RX   MEDLINE=96067599; PubMed=7487939;
RA   Boehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
RT   "Human dipeptidyl peptidase IV gene promoter: tissue-specific
RT   regulation from a TATA-less GC-rich sequence characteristic of a
RT   housekeeping gene promoter.";
RL   Biochem. J. 311:835-843(1995).
RN   [8]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Kidney;
RX   MEDLINE=93210468; PubMed=8096237;
RA   Morrison M.E., Vijayasaradhi S., Engelstein D., Albino A.P.,
RA   Houghton A.N.;
RT   "A marker for neoplastic progression of human melanocytes is a cell
RT   surface ectopeptidase.";
RL   J. Exp. Med. 177:1135-1143(1993).
CC   -!- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC       POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC       PENULTIMATE RESIDUE IS PROLINE.
CC   -!- CATALYTIC ACTIVITY: DIPEPTIDYL-POLYPEPTIDE + H(2)O = DIPEPTIDE +
CC       POLYPEPTIDE.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC       A SOLUBLE FORM.
CC   -!- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
CC       BY PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd26.htm".
CC   ---------------------------------------------------------------------
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U13735; AAB60646.1; -.
DR   EMBL; U13710; AAB60646.1; JOINED.
DR   EMBL; U13711; AAB60646.1; JOINED.
DR   EMBL; U13712; AAB60646.1; JOINED.
DR   EMBL; U13713; AAB60646.1; JOINED.
DR   EMBL; U13714; AAB60646.1; JOINED.
DR   EMBL; U13715; AAB60646.1; JOINED.
DR   EMBL; U13716; AAB60646.1; JOINED.
DR   EMBL; U13717; AAB60646.1; JOINED.
DR   EMBL; U13718; AAB60646.1; JOINED.
DR   EMBL; U13719; AAB60646.1; JOINED.
DR   EMBL; U13720; AAB60646.1; JOINED.
DR   EMBL; U13721; AAB60646.1; JOINED.
DR   EMBL; U13722; AAB60646.1; JOINED.
DR   EMBL; U13723; AAB60646.1; JOINED.
DR   EMBL; U13724; AAB60646.1; JOINED.
DR   EMBL; U13725; AAB60646.1; JOINED.
DR   EMBL; U13726; AAB60646.1; JOINED.
DR   EMBL; U13727; AAB60646.1; JOINED.
DR   EMBL; U13728; AAB60646.1; JOINED.
DR   EMBL; U13729; AAB60646.1; JOINED.
DR   EMBL; U13730; AAB60646.1; JOINED.
DR   EMBL; U13731; AAB60646.1; JOINED.
DR   EMBL; U13732; AAB60646.1; JOINED.
DR   EMBL; U13733; AAB60646.1; JOINED.
DR   EMBL; U13734; AAB60646.1; JOINED.
DR   EMBL; M74777; AAA51943.1; -.
DR   EMBL; M80536; AAA52308.1; -.
DR   EMBL; X60708; CAA43118.1; -.
DR   EMBL; S79876; AAB35614.1; -.
DR   PIR; S24313; CDHU26.
DR   MEROPS; S09.003; -.
DR   MIM; 102720; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002469; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    766       MEMBRANE FORM OF DPP (MDPP).
FT   CHAIN        39    766       SOLUBLE FORM OF DPP (SDPP)
FT                                (BY SIMILARITY).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       29    766       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    630    630       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    708    708       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    740    740       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    150    150       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    219    219       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    281    281       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    321    321       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    520    520       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    685    685       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      6      6       K -> R (IN REF. 5).
FT   CONFLICT      7      7       V -> I (IN REF. 1).
FT   CONFLICT    437    437       S -> I (IN REF. 1).
FT   CONFLICT    557    557       T -> I (IN REF. 6).
FT   CONFLICT    663    663       D -> E (IN REF. 6).
SQ   SEQUENCE   766 AA;  88278 MW;  5FB4A2C6662D6117 CRC64;
```

Query Match 7.8%; Score 272; DB 1; Length 766;
Best Local Similarity 20.4%; Pred. No. 3.1e-10;
Matches 153; Conservative 113; Mismatches 285; Indels 200; Gaps 34;

```
Qy     15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
          |: |   :|||       : | | : |||::  ||          | :   : |
Db    101 SINDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNNTQWV----- 155

Qy     66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY---------- 109
            ||| |  ||:: : :     :|:       : ::|    || |
Db    156 ----TWSPVGHKLAYVWNND-----IYVKIEPNLPSYRITWTGKEDIIYNGITDWVYEEE 206

Qy    110 ---GVSKPLWSPDGESILVTISLGEGESIDDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
             |   |||:|  :           :| |    |   |    :|
Db    207 VFSAYSALWWSPNGTFLAYA-------QFNDTEVPLIEYSFYSDESLQ---------- 247

Qy    165 RGAYAQLVLVSV-KSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYI 223
           | : | |   |:|               | :|  |: |       ||  |
Db    248 ---YPKTVRVPYPKAG-------------AVNPTVKFFV-----VNTDSLS------- 277

Qy    224 MSLESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNAT---LSKAWL-------- 272
           |: :   |:|          :|      ||       :   ||   :|  ||
Db    278 -SVTNATSIQIT------APASMLIGDHYLC------DVTWATQERISLQWLRRIQNYSV 324

Qy    273 -----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPIWTKDSQGFY-VIGTDQG 325
              ||   ||  ||     : ::          :|       | :| |   || :|  ::|
Db    325 MDICDYDESSGRWNCLVARQHIEMSTT----GWVGRFRPSEPHFTLDGNSFYKIISNEEG 380

Qy    326 STGIYYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIASVTK---------------- 366
           | |  |:     :  :   | |   :|: |  ::|::  |
Db    381 YRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYT 440

Qy    367 ----------PDRPSELYSIPLGQEEK--QL------------TGANDKFVR--EHTIS 399
                   |:|  : ||:   :| |  ||             :  ||| :|  |   :
Db    441 KVTCLSCELNPER-CQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSA 499

Qy    400 IPEEIQYATE-----DGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
           : : :|          | :::|    |  :: |  :    |||:|::: ||      | |
Db    500 LDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADTVF 559

Qy    450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
```

```
                 : : :   |   : |||   | : ::|:    |  : :| ::|  :  |
Db     560 RLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMG-F 618

Qy     507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
           :| ||: : | ||||::|: ::|  :  ||  :    :| |  :  |        :|:  :
Db     619 VDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWEYYDSV------YTERYMG 672

Qy     566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
             ||    :  | :   |    :   |::|| ||    :|: |:  ||  :| : :
Db     673 LPTPEDNLDHYRNSTVMSRAENFKQVEYLLIHGTADDNVHFQQSAQISKALVDVGVDFQA 732

Qy     624 VRFPNASHNL-SRTGHPRQRIKRLNYISSWF 653
           : : :  | : | | |        ::|   |
Db     733 MWYTDEDHGIASSTAHQHIYTHMSHFIKQCF 763


RESULT   7
DAP1_YEAST
ID   DAP1_YEAST     STANDARD;      PRT;   931 AA.
AC   P33894;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DIPEPTIDYL AMINOPEPTIDASE A (EC 3.4.14.-) (DPAP A) (YSCIV).
GN   STE13 OR YCI1 OR YOR219C OR YOR50-9.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95066382; PubMed=7975897;
RA   Anna-Arriola S.S., Herskowitz I.;
RT   "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl
RT   aminopeptidase.";
RL   Yeast 10:801-810(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB320;
RA   Flanagan C.A., Thorner J.;
RT   "STE13.";
RL   (In) Gething M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
RL   Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RL   Press, Oxford (1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=96437977; PubMed=8840505;
RA   Galisson F., Dujon B.;
RT   "Sequence and analysis of a 33 kb fragment from the right arm of
RT   chromosome XV of the yeast Saccharomyces cerevisiae.";
RL   Yeast 12:877-885(1996).
CC   -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC       ALPHA-FACTOR PRECURSOR.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC       VACUOLES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY. STRONG, TO DPAP B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L21944; AAA35119.1; -.
DR   EMBL; U08230; AAA17897.1; -.
DR   EMBL; X92441; CAA63182.1; -.
DR   EMBL; Z75127; CAA99437.1; -.
DR   PIR; A49737; A49737.
DR   PIR; S45451; S45451.
DR   MEROPS; S09.005; -.
DR   SGD; S0005745; STE13.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002469; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE_NEG.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
FT   DOMAIN        1    119       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    120    140       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN      141    931       LUMENAL (POTENTIAL).
FT   ACT_SITE    785    785       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    863    863       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    896    896       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    377    377       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   931 AA;  107200 MW;  81AF70094093C023 CRC64;


  Query Match             7.7%;  Score 269;  DB 1;  Length 931;
  Best Local Similarity   20.1%;  Pred. No. 6.4e-10;
  Matches  155;  Conservative  108;  Mismatches  298;  Indels  212;  Gaps   30;

Qy       9 DITAIVSVTDPQYAPDGTRAAYVKSQVNQEKDS--YTSNI-----------WIYETKTG 54
           | |  |::   ::  :|        |:| : |   : :|:             || :  ||
Db     220 DETFEVNLGGNRFLYEGVEFTVSTVQINYKLDKLIFGTNLESEFRHSSKGFYWIKDLNTG 279

Qy      55 GSVPWTHGEKRSTD----------PRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKL 104
             |    ||   :            :||    : |: :       |::     | |:|:
Db     280 NIEPILPPEKSDDNYELGLSKLSYAHFSPAYNYIYFVYEN-----NLFLQQVNSGVAKKV 334

Qy     105 TDIPYGVSKPL--------------------WSPDGESILVTISLGEGESIDDREKTEQ 143
           |:    || :                     |:||       :     |:||
Db     335 TE---DGSKDIFNAKPDWIYEEEVLASDQAIWWAPDDSK--AVFARFNDTSVDDIRLNRY 389

Qy     144 DSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKEL-TSHKADHGDPAFSPDGKW 202
           :    :     |    |     :  | ||::::| :  : | : |    :   :|||
Db     390 TNMNEAYLSDTKIKYPKPGFQNPQF-DLFLVNLQNGIIYSINTGGQKD----SILYNGKW 444

Qy     203 L---VFSANLTETDDASKPHDVYIMSLESG--------------------DLKQVTP-- 236
           :    |  :  ||  ||  || :  : |                       |:  : |
Db     445 ISPDTFRFEI--TDRNSKILDVKVYDIPSSQMLTVRNTNSNLFNGWIEKTKDILSIPPKP 502

Qy     237 --HRGSFGSSSFSPDGRYLALL---------------GNEKEYKNATLSKAWLYDIEQG 278
             | :|     | | : |                   || :   | :     |: |
Db     503 ELKRMDYGYIDIHADSRGFSHLFYYPTVFAKEPIQLTKGNWEVTGNGIVG----YEYETD 558

Qy     279 RLTCLTEMLDV---HLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTDQGSTGIYYISIE 335
           :     : |   ||    : ||      |      |   ||
Db     559 TIFFTANEIGVMSQHLYSISLTDSTTQNTFQSLQNPSDKYDFY---------------- 601

Qy     336 GLVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRP-----------SELYSIPLGQEE 382
                 | ||   :: |:    || |             :|::   : |
Db     602 -------------DFELSSSARYAISKKLGPDTPIKVAGPLTRVLNVAEIHDDSILQLT 647

Qy     383 KQLTGANDKF---VREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGG 439
           |     ::||   :: : : |         :||| :|   ::|| : :  ||:::||:||
Db     648 K-----DEKFKEKIKNYDLPITSYKTMVLDDGVEINYIEIKPANLNPKKKYPILVNIYGG 702

Qy     440 PHMMYGHTYF-------HEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDD 492
           |    |   |         |  |::     |: | |||: | |  | : |   |   | : |
Db     703 P----GSQTFTTKSSLAFEQAVVSGLDVIVLQIEPRGTGGKGWSFRSWAREKLGYWEPRD 758

Qy     493 VMQAVDEAIKRD-PHIDPKRLGVTGGSYGGFMTNWIVGQTN--RFKAAVTQRSISNWISF 549
           : :   : |:|:  |||  :: : | |||||  :   |   |   || |:    ::||  :
Db     759 ITEVTKKFIQRNSQHIDESKIAIWGWSYGGFTSLKTVELDNGDTFKYAMAVAPVTNWTLY 818

Qy     550 HGV----------SDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
             |          : |||          ||  ::|:                |:||
Db     819 DSVYTERYMNQPSENHEGYFEVSTIQNFKSFESLKRLF---------------IVHGTF 862
```

```
Qy     599 DDRCPIEQAEQLFIALKKMG-KETKLVRFPNASHNLSRTGHPRQRI--KRLNY 648
           ||   |:   :|   |  :|     :  ||:: |::    |  |||  ::| |
Db     863 DDNVHIQNTFRLVDQLNLLGLTNYDMHIFPDSDHSIRY--HNAQRIVFQKLYY 913


RESULT   8
DPP4_MOUSE
ID   DPP4_MOUSE     STANDARD;      PRT;   760 AA.
AC   P28843;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DIPEPTIDYL PEPTIDASE IV (EC 3.4.14.5) (DPP IV) (THYMOCYTE-ACTIVATING
DE   MOLECULE) (THAM).
GN   DPP4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS; TISSUE=Thymus;
RX   MEDLINE=92129288; PubMed=1370813;
RA   Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
RA   Pierres M.;
RT   "cDNA cloning for mouse thymocyte-activating molecule. A
RT   multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
RT   subgroup of serine proteases.";
RL   J. Biol. Chem. 267:2200-2208(1992).
RN   [2]
RP   REVISIONS.
RA   Marguet D.A.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10.A; TISSUE=Liver;
RX   MEDLINE=95092780; PubMed=7999781;
RA   Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
RT   "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
RL   Biochemistry 33:15204-15214(1994).
RN   [4]
RP   SEQUENCE OF 1-20.
RX   MEDLINE=91302787; PubMed=1712807;
RA   Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
RA   Bernard A.-M., Gorvel J.-P., Pierres M.;
RT   "Evidence that thymocyte-activating molecule is mouse CD26
RT   (dipeptidyl peptidase IV).";
RL   J. Immunol. 147:447-454(1991).
CC   -!- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC       POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC       PENULTIMATE RESIDUE IS PROLINE.
CC   -!- CATALYTIC ACTIVITY: DIPEPTIDYL-POLYPEPTIDE + H(2)O = DIPEPTIDE +
CC       POLYPEPTIDE.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC       A SOLUBLE FORM.
CC   -!- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
CC       BY PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58384; CAA41274.1; -.
DR   EMBL; U12620; AAA82213.1; -.
DR   EMBL; U12599; AAA82213.1; JOINED.
DR   EMBL; U12600; AAA82213.1; JOINED.
DR   EMBL; U12601; AAA82213.1; JOINED.
DR   EMBL; U12602; AAA82213.1; JOINED.
DR   EMBL; U12603; AAA82213.1; JOINED.
DR   EMBL; U12604; AAA82213.1; JOINED.
DR   EMBL; U12605; AAA82213.1; JOINED.
DR   EMBL; U12606; AAA82213.1; JOINED.
DR   EMBL; U12607; AAA82213.1; JOINED.
DR   EMBL; U12608; AAA82213.1; JOINED.
DR   EMBL; U12609; AAA82213.1; JOINED.
DR   EMBL; U12610; AAA82213.1; JOINED.
DR   EMBL; U12611; AAA82213.1; JOINED.
DR   EMBL; U12612; AAA82213.1; JOINED.
DR   EMBL; U12613; AAA82213.1; JOINED.
DR   EMBL; U12614; AAA82213.1; JOINED.
DR   EMBL; U12615; AAA82213.1; JOINED.
DR   EMBL; U12616; AAA82213.1; JOINED.
DR   EMBL; U12617; AAA82213.1; JOINED.
DR   EMBL; U12618; AAA82213.1; JOINED.
DR   EMBL; U12619; AAA82213.1; JOINED.
DR   PIR; S23752; S23752.
DR   MEROPS; S09.003; -.
DR   MGD; MGI:94919; Dpp4.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002469; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    760       MEMBRANE FORM OF DPP (MDPP).
FT   CHAIN        37    760       SOLUBLE FORM OF DPP (SDPP).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       29    760       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      295    466       CYS-RICH.
FT   ACT_SITE    624    624       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    702    702       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    734    734       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    213    213       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    223    223       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    514    514       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   760 AA;  87436 MW;  A5F644B46E4A3DF8 CRC64;

  Query Match             7.1%;  Score 247;  DB 1;  Length 760;
  Best Local Similarity   19.9%;  Pred. No. 1.3e-08;
  Matches  145;  Conservative  126;  Mismatches  281;  Indels  178;  Gaps   34;

Qy       2 KKLITADDI---TAIVSVTDPQYAPDGTRAA-------YVKSQVNQEKDSYTS----NI- 46
           ::||| : |   |  ::     ::|:| : |       ||| : :      ||    |:
Db     134 RQLITEEKIPNNTQWIT-----WSPEGHKLAYVWKNDIYVKVEPHLPSHRITSTGEENVI 188

Qy      47 ------WIYETKTGG--SVPWTHGEKRSTDPRWSPDGRTLAF--ISDREGDAAQLYIMST 96
                |:|| :  |  |  |           |||:   ||:   :|      :    |
Db     189 YNGITDWVYEEEVFGAYSALW----------WSPNNTFLAYAQFNDTGVPLIEYSFYSD 237

Qy      97 EGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDREKTEQDSYEPVEV----- 151
           |  :  |   |||  :  : :|  :  :| |           :   |  |:::
Db     238 ESLQYPKTVWIPYPKAGAV-NPTVKFFIVNID----------SLSSSSSAAPIQIPAPAS 286

Qy     152 --QGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANL 209
             :|  |  |    |   :  | ::: :  :  :                    ||
Db     287 VARGDHYLCDVVWATEERISLQWLRRIQNYSVMAICDYD----------------KINL 329

Qy     210 TETDDASKPHDVYIMSLESGDLKQVTPHRGSFGSSSF----SPDGRYLALLGNEKEYKNA 265
           |    : : |     :   |  :   ||  | ||| :      || |  :    |: |:
```

```
Db    330 TWNCPSEQQHVEMSTTGWVGRFRPAEPHFTSDGSSFYKIISDKDG-YKHICHFPKDKKDC 388

Qy    266 TL--SKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTD 323
          |     ||             |:: :   :|| | |
Db    389 TFITKGAW------------EVISI---EALTSDYL---------------------- 409

Qy    324 QGSTGIYYISIE-------GLVYPIRLEKEYIN----SFSLSPDE-QHFIASVTK----- 366
              |||| :          :| |:|  :: |     |  |:|:  |::  | :|
Db    410 ------YYISNQYKEMPGGRNLYKIQL-TDHTNVKCLSCDLNPERCQYYAVSFSKEAKYY 462

Qy    367 ------PDRPSELYSIPLGQEEKQLTGANDKFVREHTISIPEEIQYATE--DGVMVNG-- 416
               |  |  ||::    : |:|    |    :  :   :::|  ::  | :::|
Db    463 QLGCWGPGLP--LYTLHRSTDHKELRVLEDNSALDRML---QDVQMPSKKLDFIVLNETR 517

Qy    417 -W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYFH---EFQVLAAKGYAVVYINPRGS 470
           |  :: |    :    |||:|::: ||        |        : : :   |   : |||
Db    518 FWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADASFRLNWATYLASTENIIVASFDGRGS 577

Qy    471 HGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIVGQ 530
            | : ::|:     |  : :| ::|  : :|    :| ||: : | ||||::|: ::|
Db    578 GYQGDKIMHAINRRLGTLEVEDQIEAARQFVKMG-FVDSKRVAIWGWSYGGYVTSMVLGS 636

Qy    531 -TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAAN--V 587
           :  ||  :    :| | :  |      :|:  :   : ||    :  | :   |
Db    637 GSGVFKCGIAVAPVSRWEYYDSV------YTERYMGLPIPEDNLDHYRNSTVMSRAEHFK 690

Qy    588 ETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNL-SRTGHPRQRIKRL 646
           :  |::||  ||    :|: |:  ||   | : : : : :  | : | | |
Db    691 QVEYLLIHGTADDNVHFQQSAQISKALVDAGVDFQAMWYTDEDHGIASSTAHQHIYSHMS 750

Qy    647 NYISSWFDQH 656
          :::   |  |
Db    751 HFLQQCFSLH 760


RESULT   9
DPP4_RAT
ID   DPP4_RAT       STANDARD;      PRT;   767 AA.
AC   P14740;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DIPEPTIDYL PEPTIDASE IV (EC 3.4.14.5) (DPP IV) (GP110 GLYCOPROTEIN)
DE   (BILE CANALICULUS DOMAIN-SPECIFIC MEMBRANE GLYCOPROTEIN).
GN   DPP4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89123496; PubMed=2563382;
RA   Ogata S., Misumi Y., Ikehara Y.;
RT   "Primary structure of rat liver dipeptidyl peptidase IV deduced from
RT   its cDNA and identification of the NH2-terminal signal sequence as
RT   the membrane-anchoring domain.";
RL   J. Biol. Chem. 264:3596-3601(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88068516; PubMed=3479775;
RA   Hong W., Doyle D.;
RT   "cDNA cloning for a bile canaliculus domain-specific membrane
RT   glycoprotein of rat hepatocytes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
RN   [3]
RP   SIGNAL-ANCHOR.
RX   MEDLINE=90338089; PubMed=1974258;
RA   Hong W., Doyle D.;
RT   "Molecular dissection of the NH2-terminal signal/anchor sequence of
RT   rat dipeptidyl peptidase IV.";
RL   J. Cell Biol. 111:323-328(1990).
RN   [4]
RP   SEQUENCE OF 281-302.
RC   TISSUE=Kidney;
RX   MEDLINE=94128239; PubMed=7905271;
RA   Iwaki-Egawa S., Watanabe Y., Fujimoto Y.;
RT   "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
RT   dipeptidyl peptidase IV.";
RL   Biol. Chem. Hoppe-Seyler 374:973-975(1993).
CC   -!- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC       POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC       PENULTIMATE RESIDUE IS PROLINE.
CC   -!- CATALYTIC ACTIVITY: DIPEPTIDYL-POLYPEPTIDE + H(2)O = DIPEPTIDE +
CC       POLYPEPTIDE.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC       A SOLUBLE FORM.
CC   -!- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
CC       BY PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04591; AAA41096.1; -.
DR   EMBL; J02997; AAA41272.1; -.
DR   PIR; A33315; A33315.
DR   MEROPS; S09.003; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002469; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    767       MEMBRANE FORM OF DPP (MDPP).
FT   CHAIN        37    767       SOLUBLE FORM OF DPP (SDPP).
FT   CHAIN       281    767       60 KDA SOLUBLE FORM OF DPP.
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       29    767       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    631    631       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    709    709       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    741    741       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    217    217       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    521    521       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    686    686       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     38     38       A -> R (IN REF. 2).
FT   CONFLICT    183    183       I -> T (IN REF. 2).
FT   CONFLICT    332    332       T -> N (IN REF. 2).
FT   CONFLICT    352    352       C -> V (IN REF. 2).
FT   CONFLICT    394    394       V -> D (IN REF. 2).
FT   CONFLICT    562    562       L -> F (IN REF. 2).
FT   CONFLICT    624    624       R -> Q (IN REF. 2).
SQ   SEQUENCE   767 AA;  88003 MW;  AA17C0EC6F0E4652 CRC64;


  Query Match             7.0%;  Score 242.5;  DB 1;  Length 767;
  Best Local Similarity   18.2%;  Pred. No. 2.5e-08;
  Matches  134;  Conservative  122;  Mismatches  283;  Indels  197;  Gaps   26;

Qy      15 SVTDPQYAPDGTRAAYVKSQVNQEKDSYTSNIWIY---------ETKTGGSVPWTHGEKR 65
           |::|   :||        : | | : |||::  ||         | |   :  |
Db      99 SISDYSVSPDRLFVLLEYNYVKQWRHSYTASYSIYDLNKRQLITEEKIPNNTQWI----- 153
```

```
Qy      66 STDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIP---------------- 108
           || :|  ||::   :     :|:       : ::|
Db     154 ----TWSQEGHKLAYVWKND-----IYVKIEPHLPSHRITSTGKENVIFNGINDWVYEEE 204

Qy     109 -YGVSKPL-WSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRG 166
            :|    | |||:|  :        |  : :      :| :  :   : | : |
Db     205 IFGAYSALWWSPNGTFLAYAQFNDTGVPLIEYSFYSDESLQYPKTVWIPYPKAG------ 258

Qy     167 AYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSL 226
                                  | :|  |: : : :   :   : |
Db     259 --------------------------AVNPTVKFFIVNTDSLSSTTTTIP-------- 282

Qy     227 ESGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCL--- 283
           |:|       :| :   ||  :    | :   :|  ||  |:   :  :
Db     283 -----MQIT------APASVTTGDHYLCDVAWVSEDR---ISLQWLRRIQNYSVMAICDY 328

Qy     284 ---TEMLDVHLADALIGDSLIGGAEQ----RPIWTKDSQGFYVIGTDQG----------- 325
               | : :      |  |  |   :     | :| |   || | :|:
Db     329 DKTTLVWNCPTTQEHIETSATGWCGRFRPAEPHFTSDGSSFYKIVSDKDGYKHICQFQKD 388

Qy     326 ----------------------STGIYYISIE-------GLVYPIRL----------- 343
                                  |  :|||| |         :| |:|
Db     389 RKPEQVCTFITKGAWEVISIEALTSDYLYYISNEYKEMPGGRNLYKIQLTDHTNKKCLSC 448

Qy     344 ----EKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQL-----TGANDKFVR 394
           |:    | ||| : :::      |  |  ||::    ::|:|       | || ::
Db     449 DLNPERCQYYSVSLSKEAKYYQLGCRGPGLP--LYTLHRSTDQKELRVLEDNSALDKMLQ 506

Qy     395 EHTISIPEEIQYATEDGVMVNG---W--LMRPAQMEGETTYPLILNIHGGPHMMYGHTYF 449
           :  : :| :       | :::|     |  :: |   :     |||:::: ||        |
Db     507 D--VQMPSK----KLDFIVLNETRFWYQMILPPHFDKSKKYPLLIDVYAGPCSQKADAAF 560

Qy     450 H---EFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPH 506
                 : :  :   |   : |||   | : ::|:    |  : :| ::|  : :|
Db     561 RLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINKRLGTLEVEDQIEAARQFLKMG-F 619

Qy     507 IDPKRLGVTGGSYGGFMTNWIVGQ-TNRFKAAVTQRSISNWISFHGVSDIGYFFTDWQLE 565
           :| ||: : | ||||::|: ::|  :  ||  :    :| |  :  |      :|:  :
Db     620 VDSKRVAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWEYYDSV------YTERYMG 673

Qy     566 HDMFEDTEKLWDRSPLKYAAN--VETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKL 623
               ||    :  | :   |    :  |::||  ||    :|: |:  ||   | : :
Db     674 LPTPEDNLDHYRNSTVMSRAENFKQVEYLLIHGTADDNVHFQQSAQISKALVDAGVDFQA 733

Qy     624 VRFPNASHNL-SRTGH 638
           : : :  | : | | |
Db     734 MWYTDEDHGIASSTAH 749
```

```
RESULT  10
DPP6_RAT
ID   DPP6_RAT       STANDARD;      PRT;   859 AA.
AC   P46101;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DIPEPTIDYL PEPTIDASE IV LIKE PROTEIN (DIPEPTIDYL AMINOPEPTIDASE-
DE   RELATED PROTEIN) (DIPEPTIDYLPEPTIDASE VI) (DPPX).
GN   DPP6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92108018; PubMed=1729689;
RA   Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
RT   "Differential expression of two distinct forms of mRNA encoding
RT   members of a dipeptidyl aminopeptidase family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC       FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC       ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC       FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DPPX-L (SHOWN HERE) AND DPPX-S;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME
CC       PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS; IN
CC       CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76426; AAC42061.1; -.
DR   EMBL; M76427; AAC42062.1; -.
DR   MEROPS; S09.973; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002469; -.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT   DOMAIN        1     89       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      90    110      SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN        111   859      EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD      167   167      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      168   168      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      313   313      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      398   398      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      465   465      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      529   529      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      560   560      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      807   807      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC        1    75       MASLYQRFTGKINTSRSFPAPPEASHLLGGQGPEEDAGSKP
FT                                LGPQAQAVAPRERGGAGGRPRFQYQARSDCDEED -> MTT
FT                                AKEPSASGKSVQQQDQ (IN ISOFORM DPPX-S).
FT
SQ   SEQUENCE   859 AA;  97301 MW;  CE26856D26ED126B CRC64;
```

```
Query Match             6.7%;  Score 233;  DB 1;  Length 859;
Best Local Similarity   19.9%;  Pred. No. 1.2e-07;
Matches  147;  Conservative  104;  Mismatches  246;  Indels  242;  Gaps   32;

Qy      31 VKSQVNQEKD----SYTSNIWIYETKTG----GSVPWTHGEKRSTDP----------RW 71
           :: ::: :|:    ||        : ||      :|  ||: :| ||           |
Db     184 IRYEISPDKEYALFSYNVEPVYQHSHTGYYVLSKIP--HGDPQSLDPPEVSNAKLQYAGW 241

Qy      72 SPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTD------IPYGVSKPL---------- 115
            | |: | || :        :|  :  | :| ::       |  |:|  |
Db     242 GPKGQQLIFIFEN-----NIYYCAHVGKQAIRVVSTGKEGVIYNGLSDWLYEEEILKSHI 296

Qy     116 ---WSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLV 172
               |||||  :                |  ||  |:             :   |  |
Db     297 AHWWSPDGTRLAYA-------------TINDSRVPL------------MELPTYTGSV 329

Qy     173 LVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLK 232
             :||          ||   :|: |     |  |          ||: :|  :   ::
Db     330 YPTVKPYHYP-----KAGSENPSISLH----VIGLN-------GPTHDLEMMPPDDPRMR 373

Qy     233 QVTPHRGSFGSSSFSPDGRYLALLGNEKEYKNATLSK---AWLYDIEQGRLTCLTEMLDV 289
           :                |: ::      | || :|    ||   :  :: ||
Db     374 EY----------------YITMV------KWATSTKVAVTWLNRAQ--NVSILT----- 403

Qy     290 HLADALIG----------DSLIGGAEQRPIWTKDSQGFYVI-GTDQGSTG-IYYISI--- 334
           | ||  |          :: :    : |:::|| : |: :    ||  |  |:|::
Db     404 -LCDATTGVCTKKHEDESEAWLHRQNEEPVFSKDGRKFFFVRAIPQGGRGKFYHITVSSS 462
```

```
Qy     335 --------------------EGLVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSELY 374
           |  |  |  :  |  |  |     |  :|  :: |  |   :
Db     463 QPNSSNDNIQSITSGDWDVTEILTYDEKRNKLYFLSTEDLPRRRHLYSANTVDDFNRQCL 522

Qy     375 SIPL------------------------GQEEKQLTGANDKFVREHTISIPEEIQYATE 409
           |  |                        |     :    ||  |   :   |::| |
Db     523 SCDLVENCTYVSASFSHNMDFFLLKCEGPGVPTVTVHNTTDK-RRMFDLEANEQVQKAIY 581

Qy     410 DGVM---------VNGW-----LMRPAQMEGETTYPLILNIHGGPHMMYGHTYFH---EF 452
           |  |         |  :     :::||       |||:| : | |         |    |
Db     582 DRQMPKIEYRKIEVEDYSLPMQILKPATFTDTAHYPLLLVVDGTPGSQSVSERFEVTWET 641

Qy     453 QVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRL 512
           ::::  |  ||  : |||   | : :: ||   |  :  | |:||   :| : :||  |:
Db     642 VLVSSHGAVVVKCDGRGSGFQGTKLLHEVRRRLGFLEEKDQMEAVRTMLK-EQYIDKTRV 700

Qy     513 GVTGGSYGGFMTNWIV-------GQT-------------NRFKAAVTQRSISNWISFHGV 552
           | |  |||::: :|:       |||                : :| ::|    ::  ||:
Db     701 AVFGKDYGGYLSTYILPAKGENQGQTFTCGSALSPITDFKLYASAFSER----YLGLHGL 756

Qy     553 SDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFI 612
           :  |  |  :| |                : :|  :   ||:|   |::   :   :|
Db     757 DNRAYEMT--KLAH--------------RVSALEDQQFLIIHATADEKIHFQHTAELIT 799

Qy     613 ALKKMGKETKLVRFPNASH 631
           | |       |  :|: ||
Db     800 QLIKGKANYSLQIYPDESH 818
```

```
RESULT  11
DPP6_HUMAN
ID   DPP6_HUMAN     STANDARD;      PRT;   865 AA.
AC   P42658;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DIPEPTIDYL PEPTIDASE IV LIKE PROTEIN (DIPEPTIDYL AMINOPEPTIDASE-
DE   RELATED PROTEIN) (DIPEPTIDYLPEPTIDASE VI) (DPPX).
GN   DPP6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Hippocampus;
RX   MEDLINE=93372805; PubMed=8103397;
RA   Yokotani N., Doi K., Wenthold R.J., Wada K.;
RT   "Non-conservation of a catalytic residue in a dipeptidyl
RT   aminopeptidase IV-related protein encoded by a gene on human
RT   chromosome 7.";
RL   Hum. Mol. Genet. 2:1037-1039(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC       FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC       ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC       FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DPPX-L (SHOWN HERE) AND DPPX-S;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M96859; AAA35760.1; -.
DR   EMBL; M96860; AAA35761.1; -.
DR   MEROPS; S09.973; -.
DR   MIM; 126141; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002469; -.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT   DOMAIN          1     95       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM       96    116       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                  (POTENTIAL).
FT   DOMAIN        117    865       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD      173    173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      404    404       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      535    535       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      813    813       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC        1     81       MASLYQRFTGKINTSRSFPAPPEASHLLGGQGPEEDGGAGA
FT                                  KPLGPRAQAAAPRERGGGGGGAGGRPRFQYQGRSDGDEED
FT                                  -> MTTAKEPSASGKSVQQQEQE (IN ISOFORM DPPX-
FT                                  S).
SQ   SEQUENCE   865 AA;  97588 MW;  14B1AE0E0024464B CRC64;
```

```
 Query Match             6.4%;  Score 224;  DB 1;  Length 865;
 Best Local Similarity   20.1%;  Pred. No. 4.7e-07;
 Matches  165;  Conservative  117;  Mismatches  263;  Indels  274;  Gaps   40;

Qy       2 KKLITADDI-TAIVSVTDPQYAPDGTRAAYVK-SQVNQEKDSYTSNIWIYETKTGGSVPW 59
           || :| :|: :    : ||:       | ::  ::    :  |  :|  || |  |
Db     128 KKKVTVEDLFSEDFKIHDPE-------AKWISDTEFIYREQKGTVRLWNVETNT--STVL 178

Qy      60 THGEK----RSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPL 115
             |:|    |:     |||     |  : |     :|  |  |     |: ||:|  : |
Db     179 IEGKKIESLRAIRYEISPDREYALFSYNVE----PIYQHSYTGYYV--LSKIPHGDPQSL 232

Qy     116 --------------WSPDGESILVTISLGEGESIDDREKTEQDSYEPVEV--QGLSYKRD 159
                        | | |: ::                  | : |    |  | :
Db     233 DPPEVSNAKLQYAGWGPKGQQLIFIF-------------ENNIYYCAHVGKQAIRVVST 278

Qy     160 GK------GLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETD 213
           ||      ||:   | : :| :                  |    :||||  | ::|     :
Db     279 GKEGVIYNGLSDWLYEEEILKT--------------HIAHWWSPDGTRLAYAA----IN 319

Qy     214 DASKPHDVYIMSLE--SGDL-KQVTP-HRGSFGSSSFSPDGRYLALLG------------ 257
           |:  |    || |   :| :   | | |    || : |     : | |
Db     320 DSRVP----IMELPTYTGSIYPTVKPYHYPKAGSENPSISLHVIGLNGPTHDLEMMPPDD 375

Qy     258 -NEKEY-----KNATLSK---AWLYDIEQGRLTCLTEMLDVHLADALIG----------D 298
            :||       | || :|    ||   :   :: ||       | ||  |          :
Db     376 PRMREYYITMVKWATSTKVAVTWLNRAQ--NVSILT------LCDATTGVCTKKHEDESE 427

Qy     299 SLIGGAEQRPIWTKDSQGFY---------------------------------------- 318
           : :    : |:::|| : |:
Db     428 AWLHRQNEEPVFSKDGRKFFFIRAIPQGGRGKFYHITVSSSQPNSSNDNIQSITSGDWDV 487

Qy     319 --VIGTDQGSTGIYYISIEGL-----VYPIRLEKEYIN----------------SFSLSP 355
             ::  |:    ||::| | |     :|    |  : .                  |||  |
Db     488 TKILAYDEKGNKIYFLSTEDLPRRRQLYSANTEGNFNRQCLSCDLVENCTYFSASFSHSM 547

Qy     356 DEQHFIASVTKPDRP----------SELYSIPLGQEEKQLTGANDKFVREHTISIPE-EI 404
           |   |:      |  |           ::: :   :  |:    ||:        :|: |
Db     548 D--FFLLKCEGPGVPMVTVHNTTDKKKMFDLETNEHVKK--AINDR-------QMPKVEY 596

Qy     405 QYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMYGHTYFHEFQV------LAAK 458
           :   |   :  :::||     | |||:| : | |       :   :|:|      :::
Db     597 RDIEIDDYNLPMQILKPATFTDTTHYPLLLVVDGTPG---SQSVAEKFEVSWETVMVSSH 653

Qy     459 GYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGS 518
           |  ||  : |||   | : :: ||   |  :  | |:||   :| : :||  |: | |
Db     654 GAVVVKCDGRGSGFQGTKLLHEVRRRLGLLEEKDQMEAVRTMLK-EQYIDRTRVAVFGKD 712
```

```
Qy     519 YGGFMTNWIV-------GQT------------NRFKAAVTQRSISNWISFHGVSDIGYF 558
           |||::: :|:       |||              : :| ::|    ::  ||: :  |
Db     713 YGGYLSTYILPAKGENQGQTFTCGSALSPITDFKLYASAFSER----YLGLHGLDNRAYE 768

Qy     559 FTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMG 618
           |  :: |               : :|  |   ||:|   |::   :   :|   | :
Db     769 MT--KVAH--------------RVSALEEQQFLIIHPTADEKIHFQHTAELITQLIRGK 811

Qy     619 KETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
             | :|: ||              : ||   |||
Db     812 ANYSLQIYPDESH--------------YFTSSSLKQHL 835


RESULT  12
DAP2_YEAST
ID   DAP2_YEAST     STANDARD;      PRT;   818 AA.
AC   P18962;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DIPEPTIDYL AMINOPEPTIDASE B (EC 3.4.14.-) (DPAP B) (YSCV).
GN   DAP2 OR YHR028C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89174971; PubMed=2647766;
RA   Roberts C.J., Pohlig G., Rothman J.H., Stevens T.H.;
RT   "Structure, biosynthesis, and localization of dipeptidyl
RT   aminopeptidase B, an integral membrane glycoprotein of the yeast
RT   vacuole.";
RL   J. Cell Biol. 108:1363-1373(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC       VACUOLES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY. STRONG, TO DPAP A.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15484; CAA33512.1; -.
DR   EMBL; U10399; AAB68879.1; -.
DR   PIR; A30107; A30107.
DR   PIR; S46780; S46780.
DR   MEROPS; S09.006; -.
DR   SGD; S0001070; DAP2.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002469; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     29       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     30     45       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       46    818       LUMENAL (POTENTIAL).
FT   ACT_SITE    679    679       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    756    756       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    789    789       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    392    392       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    421    421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    738    738       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     83     83       Q -> H (IN REF. 1).
FT   CONFLICT    125    125       S -> N (IN REF. 1).
FT   CONFLICT    182    188       FEEIGNE -> LRRLET (IN REF. 1).
FT   CONFLICT    200    200       D -> N (IN REF. 1).
FT   CONFLICT    366    375       TSNVVRNESS -> DFKRGKERKF (IN REF. 1).
FT   CONFLICT    808    818       AKRAFDGQFVK -> QSVLSMGNLTNELTIYSSSHRDIHKT
FT                                FSYLHTMYI (IN REF. 1).
FT
SQ   SEQUENCE   818 AA;  93404 MW;  318F450445375BD3 CRC64;


  Query Match             6.3%;  Score 218.5;  DB 1;  Length 818;
  Best Local Similarity   19.9%;  Pred. No. 9.9e-07;
  Matches  145;  Conservative   98;  Mismatches  239;  Indels  247;  Gaps   35;

Qy      21 YAPDGTRAAYVKSQVNQEKDSYTSNIWIYET---KT-------------GGSVPWTHGEK 64
           ::|:    |||:              :||:||     ||                |   | : |:
Db     194 WSPNSNDIAYVQD----------NNIYIYSAISKKTIRAVTNDGSSFLFNGKPDWVYEEE 243

Qy      65 RSTDPR---WSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGE 121
             | :   ||| |  |||:   | :  :                        |||
Db     244 VFEDDKAAWWSPTGDYLAFLKIDESEVGEFI-------------IPYYV---------- 279

Qy     122 SILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEM 181
                          : |:| |   |:: : | : |   |   :|:| : |:| |
Db     280 -----------------QDEKDIYP--EMRSIKYPKSG---TPNPHAELWVYSMKDG-- 314

Qy     182 KELTSHKADHGDPAFSPDGKWLVFSAN-------LTETDDAS-------------KPHDV 221
             : |    |:     ||  |:           | :| | |                 | :|
Db     315 --TSFHPRISGN---KKDGSLLITEVTWVGNGNVLVKTTDRSSDILTVFLIDTIAKTSNV 369

Qy     222 YIMSLESGDLKQVT-------------PHRGSFGSSSFSPDGRYLALL----GNEKEYK 263
              :|   ::|                || |       | | |  |       |   ||
Db     370 VRNESSNGGWWEITHNTLFIPANETFDRPHNGYV---DILPIGGYNHLAYFENSNSSHYK 426

Qy     264 NATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQGFYVIGTD 323
             | |   :::  | |                    ||:    | |           | | |
Db     427 TLTEGK---WEVVNGPLAF-------------DSM----ENR---------LYFISTR 455

Qy     324 QGST--GIYYISIEG---LVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRP-------- 370
           : ||   :||| :     ::       :: :   | |   :  : :   | |
Db     456 KSSTERHVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFH 515

Qy     371 ----------------------------SELYSIPLGQEEKQLTGANDKFVREHTISIP 401
                                        | |::|  :  ::|    |:| ::
Db     516 SRKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVP-RKSFRELNLGKDEFGKD------ 568

Qy     402 EEIQYATEDGVMVNGWLMRPAQMEGETT---YPLILNIHGGPH-MMYGHTYFHEF-QVLA 456
                 ::|| : : |   : ||    ||:    :|||:       |:   | :|:|
Db     569 ----------ILVNSYEILPNDFD-ETLSDHYPVFFFAYGGPNSQQVVKTFSVGFNEVVA 617

Qy     457 AKGYA-VVYINPRGSHGYGQEFVNAVR---GDYGGKDYDDVMQAVDEAIKRDPHIDPKRL 512
           ::  | || :: ||:   ||:| : ||   |||  :|                :||:::
Db     618 SQLNAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGSL----TFVDPQKI 673

Qy     513 GVTGGSYGGFMTNWIVGQT--NRFKAAVTQRSISNWISFHGVSDIGYFFT-----DWQLE 565
             : | ||||::|   : :    ||  ::   :::|  :  |    |  |     |  :|
```

```
Db    674 SLFGWSYGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVE 733

Qy    566 HDMFEDTEKLWDRSPLKYAANVETPLLILHGERDDRCPIEQAEQLFIALKKMGKETKLVR 625
          :   |           |       |::|| ||    : : :    |   | |   |
Db    734 SSVHNVT----------ALAQANRFLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVH 782

Qy    626 -FPNASHNL 633
           ||:: |::
Db    783 VFPDSDHSI 791


RESULT  13
DPP6_BOVIN
ID   DPP6_BOVIN     STANDARD;      PRT;   863 AA.
AC   P42659;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DIPEPTIDYL PEPTIDASE IV LIKE PROTEIN (DIPEPTIDYL AMINOPEPTIDASE-
DE   RELATED PROTEIN) (DIPEPTIDYLPEPTIDASE VI) (DPPX).
GN   DPP6.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=92108018; PubMed=1729689;
RA   Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
RT   "Differential expression of two distinct forms of mRNA encoding
RT   members of a dipeptidyl aminopeptidase family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC       FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC       ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC       FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DPPX-L (SHOWN HERE) AND DPPX-S;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
CC       IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
CC       BRAIN, KIDNEY, OVARY AND TESTIS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76428; AAC41622.1; -.
DR   EMBL; M76429; AAC41623.1; -.
DR   MEROPS; S09.973; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002469; -.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT   DOMAIN        1     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN      115    863       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    171    171       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    811    811       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     79       MASLYQRFTGKINTSRSFPAPPEASRLLGGQGPEEDGAGPK
FT                                PLGAQAPAAAPRERGGGGGAGGRPRFQYQARSDCDDED ->
FT                                MTTAKEPNASGKSVQQQEQ (IN ISOFORM DPPX-S).
SQ   SEQUENCE   863 AA;  96556 MW;  23DBA792B841A39D CRC64;


  Query Match             6.1%;  Score 211.5;  DB 1;  Length 863;
  Best Local Similarity   19.5%;  Pred. No. 3e-06;
  Matches  152;  Conservative  105;  Mismatches  249;  Indels  273;  Gaps   34;

Qy     24 DGTRAAYVKSQVNQEKDSYTSNIW-IYE-TKTG----GSVPWTHGEKRSTDP-------- 69
          :  ||   :   ::|   :: |:  ||: : ||      :|  ||: :| ||
Db    183 ESLRAIRYEISPDREYALFSYNVEPIYQHSYTGYYVLSKIP--HGDPQSLDPPEVSNAKL 240

Qy     70 ---RWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTD------IPYGVSKPL----- 115
             | | |: | || :       :|  :  | :| ::        |  |:|  |
Db    241 QYAGWGPKGQQLIFIFEN-----NIYYCAHVGKQAIRVVSTGKEGVIYNGLSDWLYEEEI 295

Qy    116 --------WSPDGESILVTISLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGA 167
                  |||||                                            || |
Db    296 LKTHIAHWWSPDG-------------------------------------------TRLA 312

Qy    168 YAQLVLVSVKSGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLE 227
          || :    |   |:   |        |   |       | :|         ||:
Db    313 YATINDSRVPVMELPTYTGSVYPTAKPYHYPKAGCENPSISLHVIGLNGPTHDL------ 366

Qy    228 SGDLKQVTPHRGSFGSSSFSPDGRYLALLGNEKEY-----KNATLSKA---WLYDIEQGR 279
              ::||           ||   :      :||     | || :|    ||   :
Db    367 -----EMTP----------PDDPRM------REYYITMVKWATSTKVAVNWLSRAQ--N 402

Qy    280 LTCLTEMLDVHLADALIG----------DSLIGGAEQRPIWTKDSQGFY---------- 318
          ::||       | ||  |          :: :    : |:::|| : |:
Db    403 VSILT------LCDATTGVCTKKHEDESEAWLHRQNEEPVFSKDGRKFFFVRAIPQGGQG 456

Qy    319 ------------------------------VIGTDQGSTGIYYISIEGLVYPIRLE--- 344
                                         ::  |:  : ||::| | |  | | :
Db    457 KFYHITVSSSQPNSSNDNIQSITSGDWDVTKILSYDEKRSQIYFLSTEDL--PRRRQLYS 514

Qy    345 KEYINSF----------------SLSPDEQHFIASVTKPDRPSELYSIPLGQEEKQLT 386
             : ||                | ||    |:      |  |:   |:    ::|::
Db    515 ASTVGSFNRQCLSCDLVDNCTYFSASFSPGADFFLLKCEGPGVPT--VSVHNTTDKKKMF 572

Qy    387 GANDKFVREH---TIS---IPE-EIQYATEDGVMVNGWLMRPAQMEGETTYPLILNIHGG 439
             |    ||    ||   :|: | :    |   :   :::||       |||:| : |
Db    573 ---DLETNEHVQKAISDRQMPKVEYRKIETDDYNLPIQILKPATFTDTAHYPLLLVVDGT 629

Qy    440 PHMM-----YGHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGDYGGKDYDDVM 494
          |        :  |:  |  :::: |  ||  : |||   |   :: ||   |  :  | |
Db    630 PGSQSVAEKFAVTW--ETVMVSSHGAVVVKCDGRGSGFQGTRLLHEVRRRLGSLEEKDQM 687

Qy    495 QAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNWIV-----GQTNRFKAAVTQRSISN---- 545
          :||   :| :|:||  |: | |  |||::: :::     ||   |        |::
Db    688 EAVRVMLK-EPYIDKTRVAVFGKDYGGYLSTYLLPAKGDGQAPVFSCGSALSPITDFKLY 746

Qy    546 -------WISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLILHGER 598
                ::  ||: :  |     :: |               : :|      |::|
Db    747 ASAFSERYLGLHGLDNRAYEMA--KVAH--------------RVSALEGQQFLVIHATA 789

Qy    599 DDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNLSRTGHPRQRIKRLNYISSWFDQHL 657
          |::   :   :|   | |       |  :|: ||  |                |:   |||
Db    790 DEKIHFQHTAELITQLIKGKANYSLQIYPDESHYFS--------------SAALQQHL 833


RESULT  14
TOLB_HAEIN
ID   TOLB_HAEIN     STANDARD;      PRT;   427 AA.
AC   P44677; P94811;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   TOLB PROTEIN PRECURSOR.
GN   TOLB OR HI0382.
OS   Haemophilus influenzae.
```

```
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1479;
RX   MEDLINE=97080550; PubMed=8921895;
RA   Sen K., Sikkema D.J., Murphy T.F.;
RT   "Isolation and characterization of the Haemophilus influenzae tolQ,
RT   tolR, tolA and tolB genes.";
RL   Gene 178:75-81(1996).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE TOLB FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U32722; AAC22040.1; -.
DR   EMBL; U32470; AAC44597.1; -.
DR   TIGR; HI0382; -.
KW   Transport; Protein transport; Periplasmic; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    427       TOLB PROTEIN.
FT   VARIANT       6      6       R -> H (IN STRAIN 1479).
FT   VARIANT      14     14       V -> I (IN STRAIN 1479).
FT   VARIANT      17     19       VGS -> ITH (IN STRAIN 1479).
FT   VARIANT      21     21       A -> V (IN STRAIN 1479).
FT   VARIANT      79     79       R -> H (IN STRAIN 1479).
FT   VARIANT     129    129       T -> A (IN STRAIN 1479).
FT   VARIANT     160    160       A -> G (IN STRAIN 1479).
FT   VARIANT     237    237       A -> T (IN STRAIN 1479).
FT   VARIANT     322    322       S -> N (IN STRAIN 1479).
FT   VARIANT     326    326       A -> V (IN STRAIN 1479).
FT   VARIANT     328    328       A -> S (IN STRAIN 1479).
SQ   SEQUENCE   427 AA;  44967 MW;  0882201AEE9254B9 CRC64;
```

```
  Query Match          5.4%;  Score 190;  DB 1;  Length 427;
  Best Local Similarity  24.4%;  Pred. No. 2.8e-05;
  Matches   80;  Conservative   32;  Mismatches  102;  Indels  114;  Gaps   13;

Qy      26 TRAAYVKSQVNQEKDSYTSNIWIYETKTGGSVPWTHGEKRSTD---------------- 68
           || |||                    | ||| |:   | |  |
Db     167 TRIAYVVQ-----------------KNGGSQPY---EVRVADYDGYNQFIVNRSAQPIM 205

Qy      69 -PRWSPDGRTLAFISDREGDAAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTI 127
            | |||||: ||::|  |   :|| :     |  : :         | :|||  :
Db     206 SPAWSPDGQRLAYVS-FENKKSQLVVQDLNSGARKVVASFQGHNGAPAFSPDGSRLAFAS 264

Qy     128 SLGEGESIDDREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSH 187
           |                          |||          : ::    |   :|||
Db     265 S--------------------------RDG-------VLNIYVMGANGGTPTQLTSG 288

Qy     188 KADHGDPAFSPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGSFGSSSFS 247
           :: :||:||||  ::|      | | |    || |    |    |    |   ||:  |
Db     289 AGNNTEPAWSPDGNSILF------TSDRSGSPQVYRMDASGGSATAV----GGRGSAQIS 338

Qy     248 PDGRYLALL-GNEKEYK-------NATLSKAWL--------------YDIEQGRLTCLTE 285
            ||: | :: ||    |       :  || ::|              |   ||    | :
Db     339 ADGKTLVMINGNNNVVKQDLTTGVSEVLSTSFLGESPSLSPNGIMIIYSSTQG----LGK 394

Qy     286 MLDVHLADALIGDSLIG--GAEQRPIWT 311
           :| :  ||     || |  |  : | |:
Db     395 VLQLVSADGRFKASLPGSDGQVKFPAWS 422
```

```
RESULT  15
PPCE_PIG
ID   PPCE_PIG       STANDARD;      PRT;   710 AA.
AC   P23687;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   PROLYL ENDOPEPTIDASE (EC 3.4.21.26) (POST-PROLINE CLEAVING ENZYME)
DE   (PE).
GN   PREP.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ACTIVE SITE SER-554.
RC   TISSUE=Brain;
RX   MEDLINE=91152034; PubMed=1900195;
RA   Rennex D., Hemmings B.A., Hofsteenge J., Stone S.R.;
RT   "cDNA cloning of porcine brain prolyl endopeptidase and
RT   identification of the active-site seryl residue.";
RL   Biochemistry 30:2195-2203(1991).
RN   [2]
RP   ACTIVE SITE HIS-680.
RX   MEDLINE=91291146; PubMed=2064618;
RA   Stone S.R., Rennex D., Wikstrom P., Shaw E., Hofsteenge J.;
RT   "Inactivation of prolyl endopeptidase by a peptidylchloromethane.
RT   Kinetics of inactivation and identification of sites of
RT   modification.";
RL   Biochem. J. 276:837-840(1991).
CC   -!- FUNCTION: CLEAVES PEPTIDE BONDS ON THE C-TERMINAL SIDE OF PROLYL
CC       RESIDUES WITHIN PEPTIDES THAT ARE UP TO APPROXIMATELY 30 AMINO
CC       ACIDS LONG.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF PRO-|-XAA >> ALA-|-XAA IN
CC       OLIGOPEPTIDES.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- TISSUE SPECIFICITY: IN ALL TISSUES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9A; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M64227; AAA31110.1; -.
DR   PIR; A37942; A37942.
DR   MEROPS; S09.001; -.
DR   InterPro; IPR001375; -.
DR   InterPro; IPR002470; -.
DR   InterPro; IPR002471; -.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PRINTS; PR00862; PROLIGOPTASE.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
```

```
KW   Hydrolase; Serine protease.
FT   MOD_RES       1      1       BLOCKED.
FT   ACT_SITE    554    554       CHARGE RELAY SYSTEM.
FT   ACT_SITE    641    641       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    680    680       CHARGE RELAY SYSTEM.
FT   VARIANT      29     29       A -> H.
SQ   SEQUENCE   710 AA;  80769 MW;  70286A86238D72C0 CRC64;


Query Match             5.1%;  Score 179.5;  DB 1;  Length 710;
Best Local Similarity   19.5%;  Pred. No. 0.00027;
Matches  130;  Conservative   93;  Mismatches  238;  Indels  207;  Gaps   29;

Qy      91 LYIMSTEGGEARKLTDIPYGVSKP--------LWSPDGESILVTISLGEGESID------ 136
           ||:  :  ||||  | |  :|           :| |||    :|    : :
Db     100 LYVQDSLEGEARVFLD-PNILSDDGTVALRGYAFSEDGEYFAYGLSASGSDWVTIKFMKV 158

Qy     137 DREKTEQDSYEPVEVQGLSYKRDGKGLTRGAYAQLVLVSVKSGEMKELTSHKADHGDPAF 196
           |  |   |  | |:   :::  ||||:   || |             |:|  : :
Db     159 DGAKELPDVLERVKFSCMAWTHDGKGMFYNAYPQ-------------QDGKSDGTETST 204

Qy     197 SPDGKWLVFSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGSFGSSSFSPDGRYLALL 256
           :   |       :::|       |:  |  |  :        | :  | ||||: |
Db     205 NLHQKLYYHVLGTDQSED--------ILCAEFPDEPK------WMGGAELSDDGRYVLLS 250

Qy     257 GNEKEYKNATLSKAWLYDIEQGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIWTKDSQG 316
           |       ::: |  |::|     :| :|                 | |
Db     251 IRE---GCDPVNRLWYCDLQQ-ESNGITGILK--------------WVKLIDN 285

Qy     317 FYVIGTDQGSTGIY-YISIEGLVYPIRLEK-----EYINSFSLSPDEQHFIASVTKPDR- 369
           |         | | |:: || |:  :  :       ||     |:|  :   | : ::
Db     286 F---------EGEYDYVTNEGTVFTFKTNRHSPNYRLINIDFTDPEESKWKVLVPEHEKD 336

Qy     370 ------------------------------------------------------------ 369

Db     337 VLEWVACVRSNFLVLCYLHDVKNTLQLHDLATGALLKIFPLEVGSVVGYSGQKKDTEIFY 396

Qy     370 -------PSELYSIPLGQEEKQLTGANDKFVREHTISIPE-------EIQYATEDGVMVN 415
                |  :|   | :|| :      :  || |:   :        :| | ::||  :
Db     397 QFTSFLSPGIIYHCDLTKEELE-----PRVFREVTVKGIDASDYQTVQIFYPSKDGTKIP 451

Qy     416 GWLMRPAQMEGETTYPLILNIHGG------PHMMYGHTYF--HEFQVLAAKGYAVVYINP 467
           :::    :: : ::|  |  :||      |:       |  |   |||         |
Db     452 MFIVHKKGIKLDGSHPAFLYGYGGFNISITPNYSVSRLIFVRHMGGVLAV-------ANI 504

Qy     468 RGSHGYGQEFVNAVRGDYGGKD--YDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTN 525
           ||   ||: :     |    |   :||   | :  || : :  |||| : ||| || :
Db     505 RGGGEYGETWHKG--GILANKQNCFDDFQCAAEYLIK-EGYTSPKRLTINGGSNGGLLVA 561

Qy     526 WIVGQ-TNRFKAAVTQRSISNWISFH-----------GVSDIGYFFTDWQLEHDMFEDT 572
           |  : |   : |  : : : ||             | ||    | :| :::    :
Db     562 TCANQRPDLFGCVIAQVGVMDMLKFHKYTIGHAWTTDYGCSDSKQHF-EWLIKYSPLHNV 620

Qy     573 EKLWDRSPLKYAANVETPLLILHGERDDR-CPIEQAEQLFIA-LKKMGKETKLVRFPNAS 630
           || :   ::| :     :|:|  : |||  |:   :  ||| |: :   ::    |
Db     621 -KLPEADDIQYPS-----MLLLTADHDDRVVPLHSLK--FIATLQYIVGRSRKQNNPLLI 672

Qy     631 HNLSRTGH 638
           |  :: ||
Db     673 HVDTKAGH 680


Search completed: September 25, 2001, 17:38:05
Job time: 142 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 03:01:08 ; Search time 3698.58 Seconds
(without alignments)
8242.875 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 1971
Sequence: 1 atgaaaaagctgataaccgc.....catggtttgatcaacatctc 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1971	100.0	1971	9	AX006271	AX006271 Sequence
2	c	1971	100.0	217420	2	BSUB0017	Z99120 Bacillus su
3	c	238.2	12.1	265118	2	CNSPAX06	AJ248288 Pyrococcu
4	c	238.2	12.1	265118	9	AX041922	AX041922 Sequence
5		227.2	11.5	1896	10	E16634	E16634 Pyrococcus
6		227.2	11.5	233000	2	AP000003	AP000003 Pyrococcu
7	c	191.2	9.7	10256	1	AE001879	AE001879 Deinococc
8		161.8	8.2	251700	2	AP000062	AP000062 Aeropyrum

c	9	158	8.0	304290	2	CNSPAX05	AJ248287 Pyrococcu
c	10	158	8.0	349980	9	AX041921	AX041921 Sequence
	11	150.6	7.6	257000	2	AP000002	AP000002 Pyrococcu
	12	149.8	7.6	1869	9	AR086952	AR086952 Sequence
	13	149.8	7.6	1869	9	AR095647	AR095647 Sequence
	14	132.6	6.7	343800	2	AP000993	AP000993 Thermopla
	15	130.4	6.6	9825	1	AE005114	AE005114 Halobacte
c	16	130.4	6.6	247695	2	AP000064	AP000064 Aeropyrum
c	17	75.2	3.8	139203	73	AC068311	AC068311 Homo sapi
	18	65	3.3	1998	5	AF059278	AF059278 Entamoeba
	19	64.8	3.3	2400	2	BSTHRBC	X04603 B. subtilis
	20	62.8	3.2	1992	5	AF059279	AF059279 Entamoeba
c	21	54.4	2.8	7218	10	I66494	I66494 Sequence 14
c	22	46.4	2.4	67965	3	SSOLP2N09	AL512970 Sulfolobu
c	23	46.2	2.3	316900	3	TACID3	AL445065 Thermopla
	24	46	2.3	2325	12	AF082514	AF082514 Trichophy
c	25	45.4	2.3	13928	1	AE004038	AE004038 Xylella f
	26	41.6	2.1	3406	2	AF299085	AF299085 Myxococcu
	27	41.2	2.1	38000	15	SPBC1711	AL353012 S.pombe c
	28	39	2.0	6741	10	E32987	E32987 Gene encodi
	29	39	2.0	11452	1	AE006030	AE006030 Caulobact
c	30	39	2.0	136257	88	AC073864	AC073864 Homo sapi
	31	39	2.0	170075	70	AC026678	AC026678 Homo sapi
c	32	39	2.0	205400	82	AP000852	AP000852 Homo sapi
c	33	39	2.0	313462	69	AC024935	AC024935 Homo sapi
	34	38.4	1.9	12212	1	AE005872	AE005872 Caulobact
c	35	38.4	1.9	57223	77	AC090315	AC090315 Homo sapi
c	36	38.4	1.9	175035	69	AC025602	AC025602 Homo sapi
c	37	38	1.9	119844	61	AC010044	AC010044 Drosophil
	38	37.6	1.9	40824	6	CELF01F1	U13070 Caenorhabdi
	39	37.6	1.9	162454	66	AC021568	AC021568 Homo sapi
c	40	37.6	1.9	166393	77	AC087878	AC087878 Homo sapi
c	41	37.6	1.9	191637	62	AC011710	AC011710 Homo sapi
	42	37.6	1.9	204879	65	AC018475	AC018475 Homo sapi
	43	37.4	1.9	2196	10	E02747	E02747 DNA encodin
	44	37.4	1.9	2385	7	PIGAAREL	D00524 Sus scrofa
c	45	37.4	1.9	169950	76	AC084049	AC084049 Homo sapi

ALIGNMENTS

```
RESULT   1
AX006271
LOCUS       AX006271      1971 bp    DNA             PAT       24-AUG-2000
DEFINITION  Sequence 1 from Patent WO9903984.
ACCESSION   AX006271
VERSION     AX006271.1  GI:9929124
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 1971)
  AUTHORS   Estell,D.A.
  TITLE     Proteases from gram-positive organisms
  JOURNAL   Patent: WO 9903984-A 1 28-JAN-1999;
            GENENCOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)
FEATURES             Location/Qualifiers
     source          1. .1971
                     /organism="Bacillus subtilis"
                     /db_xref="taxon:1423"
BASE COUNT      558 a    431 c    524 g    458 t
ORIGIN


  Query Match             100.0%;  Score 1971;  DB 9;  Length 1971;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1971;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaaaaagctgataaccgcagacgacatcacagcgattgtctctgtgaccgatcctcaa 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAAAGCTGATAACCGCAGACGACATCACAGCGATTGTCTCTGTGACCGATCCTCAA 60

Qy      61 tacgccccagacggtacccgtgccgcatatgtaaaatcacaagtaaatcaagagaaagat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TACGCCCCAGACGGTACCCGTGCCGCATATGTAAAATCACAAGTAAATCAAGAGAAAGAT 120

Qy     121 tcgtatacatcaaatatatggatctatgaaacgaaaacgggaggatctgttccttggaca 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCGTATACATCAAATATATGGATCTATGAAACGAAAACGGGAGGATCTGTTCCTTGGACA 180

Qy     181 catggagaaaagcgaagcaccgacccaagatggtctccggacgggcgcacgcttgccttt 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CATGGAGAAAAGCGAAGCACCGACCCAAGATGGTCTCCGGACGGGCGCACGCTTGCCTTT 240

Qy     241 atttctgatcgagaaggcgatgcggcacagctttatatcatgagcactgaaggcggagaa 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATTTCTGATCGAGAAGGCGATGCGGCACAGCTTTATATCATGAGCACTGAAGGCGGAGAA 300

Qy     301 gcaagaaaactgactgatatcccatatggcgtgtcaaagccgctatggtccccggacggt 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCAAGAAAACTGACTGATATCCCATATGGCGTGTCAAAGCCGCTATGGTCCCCGGACGGT 360

Qy     361 gaatcgattctggtcactatcagtttgggagaggggggaaagcattgatgaccgagaaaaa 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GAATCGATTCTGGTCACTATCAGTTTGGGAGAGGGGGAAAGCATTGATGACCGAGAAAAA 420

Qy     421 acagagcaggacagctatgaacctgttgaagtgcaaggcctctcctacaaacgggacggc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ACAGAGCAGGACAGCTATGAACCTGTTGAAGTGCAAGGCCTCTCCTACAAACGGGACGGC 480

Qy     481 aaagggctgacgagaggtgcgtatgcccagcttgtgcttgtcagcgtaaagtcgggtgag 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AAAGGGCTGACGAGAGGTGCGTATGCCCAGCTTGTGCTTGTCAGCGTAAAGTCGGGTGAG 540

Qy     541 atgaaagagctgacaagtcacaaagctgatcatggtgatcctgctttttctcctgacggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATGAAAGAGCTGACAAGTCACAAAGCTGATCATGGTGATCCTGCTTTTTCTCCTGACGGC 600

Qy     601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AAATGGCTTGTTTTCTCAGCTAATTTAACTGAAACAGATGATGCCAGCAAGCCGCATGAT 660

Qy     661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GTTTACATAATGTCACTGGAGTCTGGAGATCTTAAGCAGGTTACACCTCATCGCGGCTCA 720

Qy     721 ttcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TTCGGATCAAGCTCATTTTCACCAGACGGAAGGTATCTTGCTTTGCTTGGAAATGAAAAG 780

Qy     781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GAATATAAGAATGCTACGCTCTCAAAGGCGTGGCTCTATGATATCGAACAAGGCCGCCTC 840

Qy     841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ACATGTCTTACTGAGATGCTGGACGTTCATTTAGCGGATGCGCTGATTGGAGATTCATTG 900

Qy     901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ATCGGTGGTGCTGAACAGCGCCCGATTTGGACAAAGGACAGCCAAGGGTTTTATGTCATC 960

Qy     961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGCACAGATCAAGGCAGTACGGGCATCTATTATATTTCGATTGAAGGCCTTGTGTATCCG 1020

Qy    1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ATTCGTCTGGAAAAAGAGTACATCAATAGCTTTTCTCTTTCACCTGATGAACAGCACTTT 1080

Qy    1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ATTGCCAGTGTGACAAAGCCGGACAGACCGAGTGAGCTTTACAGTATCCCGCTTGGACAG 1140
```

```
Qy    1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GAAGAGAAACAGCTGACTGGCGCGAATGACAAGTTTGTCAGGGAGCATACGATATCAATA 1200

Qy    1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CCTGAAGAGATTCAATATGCTACAGAAGACGGCGTGATGGTGAACGGCTGGCTGATGAGG 1260

Qy    1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CCTGCACAAATGGAAGGTGAGACAACATATCCACTTATTCTTAACATACACGGCGGTCCG 1320

Qy    1321 catatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggatac 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CATATGATGTACGGACATACATATTTTCATGAGTTTCAGGTGCTGGCGGCGAAAGGATAC 1380

Qy    1381 gcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatgcg 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GCGGTCGTTTATATCAATCCGAGAGGAAGCCACGGCTACGGGCAGGAATTTGTGAATGCG 1440

Qy    1441 gtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgaggct 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GTCAGAGGAGATTATGGGGGAAAGGATTATGACGATGTGATGCAGGCTGTGGATGAGGCT 1500

Qy    1501 atcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacgga 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 ATCAAACGAGATCCGCATATTGATCCTAAGCGGCTCGGTGTCACGGGCGGAAGCTACGGA 1560

Qy    1561 ggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttacccag 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GGTTTTATGACCAACTGGATCGTCGGGCAGACGAACCGCTTTAAAGCTGCCGTTACCCAG 1620

Qy    1621 cgctcgatatcaaattggatcagctttcacggcgtcagtgatatcggctatttctttaca 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 CGCTCGATATCAAATTGGATCAGCTTTCACGGCGTCAGTGATATCGGCTATTTCTTTACA 1680

Qy    1681 gactggcagcttgagcatgacatgtttgaggacacagaaaagctctgggaccggtctcct 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GACTGGCAGCTTGAGCATGACATGTTTGAGGACACAGAAAAGCTCTGGGACCGGTCTCCT 1740

Qy    1741 ttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatgac 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 TTAAAATACGCAGCAAACGTGGAGACACCGCTTTTGATACTGCATGGCGAGCGGGATGAC 1800

Qy    1801 cgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaaggaa 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 CGATGCCCGATCGAGCAGGCGGAGCAGCTGTTTATCGCTCTGAAAAAAATGGGCAAGGAA 1860

Qy    1861 accaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaaga 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 ACCAAGCTTGTCCGTTTTCCGAATGCATCGCACAATTTATCACGCACCGGACACCCAAGA 1920

Qy    1921 cagcggatcaagcgcctgaattatatcagctcatggtttgatcaacatctc 1971
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 CAGCGGATCAAGCGCCTGAATTATATCAGCTCATGGTTTGATCAACATCTC 1971
```

```
RESULT   2
BSUB0017/c
LOCUS       BSUB0017   217420 bp    DNA             BCT       26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 17 of 21): from 3197001
            to 3414420.
ACCESSION   Z99120 AL009126
VERSION     Z99120.1  GI:2635613
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 217420)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 217420)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1. .217420
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            complement(3. .1277)
                     /gene="yulE"
     CDS             complement(3. .1277)
                     /gene="yulE"
                     /function="unknown"
                     /note="similar to L-rhamnose isomerase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15107.1"
                     /db_xref="GI:2635614"
                     /db_xref="SPTREMBL:O05264"
                     /translation="MTIKANYDSAKQAYEKWGIDVEEALRQLEQVPISIHCWQGDDIE
                     GFEVNKGELSGGIDVTGNYPGKAQTPEELRRDLEKALSLIPGKHRVNLHAIYAETNRE
                     AVERDELKPQHFENWVKWAKNLGLGLDFNPTLFSHEKAADGLTLSHPDPDIREFWIRH
                     CIACRRIGEYFGKELGTPCLTNIWIPDGYKDIPSDRLTPRKRLKESLDRIFSEEISEQ
                     HNLDSIESKLFGLGSESYVVGSHEFYLAYALTNHKLCLLDTGHFHPTETVSNKISSML
                     LYTDKLALHVSRPVRWDSDHVVVLDDELREIALEIVRNHALEKVAIGLDFFDASINRV
                     AAWTIGTRNMIKALLYALLLPNGYLKQLQEEGRYTERLALMEEFKTYPFGAIWDSYCE
                     QMGVPVKEAWLYDIKEYEQQVLLKRKASSPIV"
     gene            complement(1303. .1617)
                     /gene="yulD"
     CDS             complement(1303. .1617)
                     /gene="yulD"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15108.1"
                     /db_xref="GI:2635615"
```

```
                /db_xref="SPTREMBL:O05263"
                /translation="MKRKASIMFVHQDKYEEYKQRHDDIWPEMAEALKAHGAHHYSIF
                LDEETGRLFAYLEIEDEEKWRKMADTEVCQRWWKSMAPLMKTNSDFSPVAIDLKEVFY
                LD"
gene            complement(1635. .3092)
                /gene="yulC"
CDS             complement(1635. .3092)
                /gene="yulC"
                /function="unknown"
                /note="similar to rhamnulokinase"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15109.1"
                /db_xref="GI:2635616"
                /db_xref="SPTREMBL:O05262"
                /translation="MIYTAIDVGASSGRIMVGELNEGKLDIQEIHRFANGFSQRDGHC
                LWDIDHLLKQILQGLQKVKTLGYEHCTVGIDTWAVDYVLLDEKGDRLREAISYRDRRT
                DHTIDKLEHTLSKAAIYQKTGIQFQPFNTIYQLFEEDRELLKKTDKIMMIPDYLGYCL
                TGKAVTEITNVSTTQLLNVSTGNLDPELLEAVSVLEQQFAPLTEPGCELGKLRNEWFP
                DYDLPACKVMTVATHDTASAVIAAPGVNDGWAYISSGTWSLIGVENKTPIITDLALEN
                NYTNERGANNTIRFLKNIIGMWVIQEVKQQLQADYSFQQLAEEAKKTEPFQQFINLND
                KRFLNPENMIKEIQHYCRQTRQKIPRTAGELACCIYSNLAIIYAIAIKELETITEKPI
                EQFHIIGGGARNDFLNQLTADMSGKAVYAGPIEATATGNLLMQMIAAKEVKDIKEARQ
                VVRNSFPIKVFTPKDIDRSTIIQSFQQTVLKALSK"
gene            complement(3097. .3873)
                /gene="yulB"
CDS             complement(3097. .3873)
                /gene="yulB"
                /function="unknown"
                /note="similar to transcriptional regulator (DeoR family)"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15110.1"
                /db_xref="GI:2635617"
                /db_xref="SWISS-PROT:O05261"
                /translation="MLVAERQQKIVEIVNMRSSIRVSELSDIFSVTEETIRRDLEKLE
                KEHKLSRSHGGAVSIQQKESEIHFSEREITNVIEKKAIAHEAAKYVKSGDRIILDAST
                TAWYMAKILPDIELTVITNSMKAAIELSNKENISVISTGGILLEKSLSFAGPLAERSL
                ETYHVNKTFLSCKGFDVHNGMSDSNEWQALLKKRMIERSDQTILMADSSKWGNREFSH
                IASLQDVSRLITDSGLDPASVKALEDKKVKVTAVPLSKRG"
gene            complement(3930. .5999)
                /gene="yuxG"
CDS             complement(3930. .5999)
                /gene="yuxG"
                /function="unknown"
                /note="alternate gene name: yulA
                similar to sorbitol-6-phosphate 2-dehydrogenase"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15111.1"
                /db_xref="GI:2635618"
                /db_xref="SWISS-PROT:P40747"
                /translation="MVKHIWDSERAAQLPKGVEELVYRSNLIGSDRTVCNWGGGNTSM
                KTTEKDFRGREIEVMWVKGSGSDLATMKAHNFSGLKLDDIRPLIKRDQMPDEEMVDYL
                SHCMIDSKHPRPSIETLLHAFLPYKHVDHTHPDAIISICCADNGKQIAEDIYGNRFVW
                VPYVRPGFTLSKMIAEGVANNPHAELVLMEKHGLVTWGETSETCYQKTISIIQEAEQY
                INDRINQHEVFGGKRYQPLPEDKRKQILAGIMPVIRGAVSEEKKMILSYDDHDDVLEF
                VNSVQAPALSQIGAACPDHLVHTKRVPLYIDWNPETQDVHKLADLIKSGVETFTSEYQ
                AYFTRNQQDGDQIFESAPRVILIPGIGMVNTGKSYAMSKVSGALYRRAIAVMKGATAL
                GQFVSLHENESYHVEYWPLELYKLTLAPPEAEFSRKVALITGGAGGIGSAACRRFAAE
                GGHVIVADLNIEGAQKIAGEINDAYGKGRAMAVKMDVTKEEDVQSAFERAALAYGGID
                IVVNNAGLATSSPFDETSLKEWNLNMNVLGTGYFLVAREAFKQMKHQNRGGSMVFVGS
                KNSVYAGKNASAYSSVKALETHLARCIAAEGGEFGIRVNSVLPDAVLQGSAIWGSSWR
                EERAAAYGIEPDQLEEHYRKRTALLVNIYPEDIAEAIAFFASSKAEKTTGCMITVDGG
                VPAAFTR"
terminator      complement(6109. .6124)
                /gene="tlpB"
gene            complement(6109. .8125)
                /gene="tlpB"
CDS             complement(6137. .8125)
                /gene="tlpB"
                /codon_start=1
                /transl_table=11
                /product="methyl-accepting chemotaxis protein"
                /protein_id="CAB15112.1"
                /db_xref="GI:2635619"
                /db_xref="SWISS-PROT:P39217"
                /translation="MGKFIQWIKQPSISKPLIAAFLAVLILPVGVLAYFSYQSAWNAL
                DRELISSAKGNVEELNSTLQNKLEDKVKAIDYYSETVDKDILLGKNKTLLKEKFKQYT
                TLNDDVGAIYAASEDKKLYKYPDSGVPKGFDPTGRDWYKQAVAEKGQAVFSEPYTDEA
                TGDIVVTISKQLKDGSGVIALDLNLDEVLTASKRIKIGKEGFAFITTGNKKYIAHPTI
                KPGTTGSGDWTNQVYSKKEGSFEYTFEGKEKKMAFTTNKLTGWKIAGTYFVSELQDAS
                SPVLNTAVIILCVSIVIGGILILYIIRAITKPLRKLVSTSAKISSGDLTEVIDIHSKN
                EFGQLGESFNEMSASLRSVIGVIQTSVENVASSSEELTASAAQTSKATEHITLAIEQF
                SDGNEAQSEKLETSSNHLSQMNEGISKVAQASSTITKSSIQSSEAAGSGEKLVEHTVG
                QMKTIDQSVQKAEAVVKGLETKSQDITSILNVINGIADQTNLLALNAAIEAARAGEYG
                RGFSVVAEEVRKLAVQSADSAKEIEGLIQEIVREISTSLSMFQSVNHEVKEGLQITDQ
                TAESFKQIYEMTTQISGELQNLNATVEQLSAGSQEVSSAVEDISAVAKESSAGIQDIA
                ASAEEQLASMEEISSSAETLANMAEELQDITKKFKIES"
gene            complement(8204. .10224)
                /gene="mcpA"
terminator      complement(8204. .8228)
                /gene="mcpA"
CDS             complement(8239. .10224)
                /gene="mcpA"
                /function="chemotaxis deficiency toward glucose and
                alpha-methyl-glucoside"
                /codon_start=1
                /transl_table=11
                /product="methyl-accepting chemotaxis protein"
                /protein_id="CAB15113.1"
                /db_xref="GI:2635620"
                /db_xref="SWISS-PROT:P39214"
                /translation="MKKILQLIKQRSITRKLLVSFLSILIIPVVILAIFAYQSASSSL
                DRQMMGSALENVQQLNEIINTSIGEKENSADYFSEWLTKEKYNAKSNASIREKFSQYI
                SINKDVESIYTSDTKGHFTRYPDLPMPSGYNPVERDWYKKAVANKGKVVITDPYKTAS
                TNTMVVTIAQQTKDGSGVIAINMTIENLLKTTKKVNIGTQGYAFIMTKDKKVVAHPNE
                QSGTELKGDWLDKMLSADKGDFQYTMDGDKKKMAFDTNKLTGWKIGGTMYLDEIHEAA
                QPVLHLALIVLAAAIIIGIIVMTLIIRSITTPLKQLVGSSKRISEGDLTETIDIRSKD
                ELGELGKSFNNMASSLRSLIHAIQDSVDNVAASSEELTASAAQTSKATEHITLAIEQF
                SNGNEKQNENIETAAEHIYQMNDGLTNMAQASEVITDSSVQSTEIASEGGKLVHQTVG
                QMNVIDKSVKEAEQVVRGLETKSKDITNILRVINGIADQTNLLALNAAIEAARAGEYG
                RGFSVVAEEVRKLAVQSADSAKEIEGLIIEIVKEINTSLGMFQSVNQEVQTGLDITDK
```

```
 Query Match          100.0%;  Score 1971;  DB 2;  Length 217420;
 Best Local Similarity  100.0%;  Pred. No. 0;
 Matches 1971;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgaaaaagctgataaccgcagacgacatcacagcgattgtctctgtgaccgatcctcaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114794 ATGAAAAAGCTGATAACCGCAGACGACATCACAGCGATTGTCTCTGTGACCGATCCTCAA 114735

Qy     61 tacgccccagacggtacccgtgccgcatatgtaaaatcacaagtaaatcaagagaaagat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114734 TACGCCCCAGACGGTACCCGTGCCGCATATGTAAAATCACAAGTAAATCAAGAGAAAGAT 114675

Qy    121 tcgtatacatcaaatatatggatctatgaaacgaaaacgggaggatctgttccttggaca 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114674 TCGTATACATCAAATATATGGATCTATGAAACGAAAACGGGAGGATCTGTTCCTTGGACA 114615

Qy    181 catggagaaaagcgaagcaccgacccaagatggtctccggacgggcgcacgcttgccttt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114614 CATGGAGAAAAGCGAAGCACCGACCCAAGATGGTCTCCGGACGGGCGCACGCTTGCCTTT 114555

Qy    241 atttctgatcgagaaggcgatgcggcacagctttatatcatgagcactgaaggcggagaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114554 ATTTCTGATCGAGAAGGCGATGCGGCACAGCTTTATATCATGAGCACTGAAGGCGGAGAA 114495

Qy    301 gcaagaaaactgactgatatcccatatggcgtgtcaaagccgctatggtccccggacggt 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114494 GCAAGAAAACTGACTGATATCCCATATGGCGTGTCAAAGCCGCTATGGTCCCCGGACGGT 114435

Qy    361 gaatcgattctggtcactatcagtttgggagaggggggaaagcattgatgaccgagaaaaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114434 GAATCGATTCTGGTCACTATCAGTTTGGGAGAGGGGGAAAGCATTGATGACCGAGAAAAA 114375
```

```
Qy    421 acagagcaggacagctatgaacctgttgaagtgcaaggcctctcctacaaacgggacggc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114374 ACAGAGCAGGACAGCTATGAACCTGTTGAAGTGCAAGGCCTCTCCTACAAACGGGACGGC 114315

Qy    481 aaagggctgacgagaggtgcgtatgcccagcttgtgcttgtcagcgtaaagtcgggtgag 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114314 AAAGGGCTGACGAGAGGTGCGTATGCCCAGCTTGTGCTTGTCAGCGTAAAGTCGGGTGAG 114255

Qy    541 atgaaagagctgacaagtcacaaagctgatcatggtgatcctgctttttctcctgacggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114254 ATGAAAGAGCTGACAAGTCACAAAGCTGATCATGGTGATCCTGCTTTTTCTCCTGACGGC 114195

Qy    601 aaatggcttgttttctcagctaatttaactgaaacagatgatgccagcaagccgcatgat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114194 AAATGGCTTGTTTTCTCAGCTAATTTAACTGAAACAGATGATGCCAGCAAGCCGCATGAT 114135

Qy    661 gtttacataatgtcactggagtctggagatcttaagcaggttacacctcatcgcggctca 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114134 GTTTACATAATGTCACTGGAGTCTGGAGATCTTAAGCAGGTTACACCTCATCGCGGCTCA 114075

Qy    721 ttcggatcaagctcattttcaccagacggaaggtatcttgctttgcttggaaatgaaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114074 TTCGGATCAAGCTCATTTTCACCAGACGGAAGGTATCTTGCTTTGCTTGGAAATGAAAAG 114015

Qy    781 gaatataagaatgctacgctctcaaaggcgtggctctatgatatcgaacaaggccgcctc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114014 GAATATAAGAATGCTACGCTCTCAAAGGCGTGGCTCTATGATATCGAACAAGGCCGCCTC 113955

Qy    841 acatgtcttactgagatgctggacgttcatttagcggatgcgctgattggagattcattg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113954 ACATGTCTTACTGAGATGCTGGACGTTCATTTAGCGGATGCGCTGATTGGAGATTCATTG 113895

Qy    901 atcggtggtgctgaacagcgcccgatttggacaaaggacagccaagggttttatgtcatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113894 ATCGGTGGTGCTGAACAGCGCCCGATTTGGACAAAGGACAGCCAAGGGTTTTATGTCATC 113835

Qy    961 ggcacagatcaaggcagtacgggcatctattatatttcgattgaaggccttgtgtatccg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113834 GGCACAGATCAAGGCAGTACGGGCATCTATTATATTTCGATTGAAGGCCTTGTGTATCCG 113775

Qy   1021 attcgtctggaaaaagagtacatcaatagcttttctctttcacctgatgaacagcacttt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113774 ATTCGTCTGGAAAAAGAGTACATCAATAGCTTTTCTCTTTCACCTGATGAACAGCACTTT 113715

Qy   1081 attgccagtgtgacaaagccggacagaccgagtgagctttacagtatcccgcttggacag 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113714 ATTGCCAGTGTGACAAAGCCGGACAGACCGAGTGAGCTTTACAGTATCCCGCTTGGACAG 113655

Qy   1141 gaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaata 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113654 GAAGAGAAACAGCTGACTGGCGCGAATGACAAGTTTGTCAGGGAGCATACGATATCAATA 113595

Qy   1201 cctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatgagg 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113594 CCTGAAGAGATTCAATATGCTACAGAAGACGGCGTGATGGTGAACGGCTGGCTGATGAGG 113535

Qy   1261 cctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtccg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113534 CCTGCACAAATGGAAGGTGAGACAACATATCCACTTATTCTTAACATACACGGCGGTCCG 113475

Qy   1321 catatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggatac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113474 CATATGATGTACGGACATACATATTTTCATGAGTTTCAGGTGCTGGCGGCGAAAGGATAC 113415

Qy   1381 gcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatgcg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113414 GCGGTCGTTTATATCAATCCGAGAGGAAGCCACGGCTACGGGCAGGAATTTGTGAATGCG 113355

Qy   1441 gtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgaggct 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113354 GTCAGAGGAGATTATGGGGGAAAGGATTATGACGATGTGATGCAGGCTGTGGATGAGGCT 113295

Qy   1501 atcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacgga 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113294 ATCAAACGAGATCCGCATATTGATCCTAAGCGGCTCGGTGTCACGGGCGGAAGCTACGGA 113235

Qy   1561 ggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttacccag 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113234 GGTTTTATGACCAACTGGATCGTCGGGCAGACGAACCGCTTTAAAGCTGCCGTTACCCAG 113175

Qy   1621 cgctcgatatcaaattggatcagctttcacggcgtcagtgatatcggctatttctttaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113174 CGCTCGATATCAAATTGGATCAGCTTTCACGGCGTCAGTGATATCGGCTATTTCTTTACA 113115

Qy   1681 gactggcagcttgagcatgacatgtttgaggacacagaaaagctctgggaccggtctcct 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113114 GACTGGCAGCTTGAGCATGACATGTTTGAGGACACAGAAAAGCTCTGGGACCGGTCTCCT 113055

Qy   1741 ttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatgac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113054 TTAAAATACGCAGCAAACGTGGAGACACCGCTTTTGATACTGCATGGCGAGCGGGATGAC 112995

Qy   1801 cgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaaggaa 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112994 CGATGCCCGATCGAGCAGGCGGAGCAGCTGTTTATCGCTCTGAAAAAAATGGGCAAGGAA 112935

Qy   1861 accaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaaga 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112934 ACCAAGCTTGTCCGTTTTCCGAATGCATCGCACAATTTATCACGCACCGGACACCCAAGA 112875

Qy   1921 cagcggatcaagcgcctgaattatatcagctcatggtttgatcaacatctc 1971
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112874 CAGCGGATCAAGCGCCTGAATTATATCAGCTCATGGTTTGATCAACATCTC 112824
```

```
RESULT    3
CNSPAX06/c
LOCUS       CNSPAX06    265118 bp    DNA              BCT       09-SEP-1999
DEFINITION  Pyrococcus abyssi complete genome; segment 6/6.
ACCESSION   AJ248288 AL096836
VERSION     AJ248288.1  GI:5458960
KEYWORDS    complete genome.
SOURCE      Pyrococcus abyssi.
  ORGANISM  Pyrococcus abyssi
            Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
            Pyrococcus.
REFERENCE   1  (bases 1 to 265118)
  AUTHORS   Heilig,R.
  TITLE     Pyrococcus abyssi genome sequence: insights into archaeal
            chromosome structure and evolution
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 265118)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .265118
                     /organism="Pyrococcus abyssi"
                     /strain="Orsay"
                     /db_xref="taxon:29292"
     gene            164. .1615
                     /gene="bgal"
     CDS             164. .1615
                     /gene="bgal"
                     /note="PAB2376"
                     /codon_start=1
                     /transl_table=11
                     /product="BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE)."
                     /protein_id="CAB50447.1"
                     /db_xref="GI:5458961"
                     /translation="MEFYWGVVQSAFQFEMGDPYRRNLDTRSDWWHWVRDPYNIKNEL
                     VSGDLPEEGINNYELYEIDHRLAKELGLNAYQLTIEWSRIFPCPTYSVEVEVERDGYG
                     FIKSVKIRKEHLEKLDELANKREVRHYENVLKNLKKLNFTTFVTLNHQTNPIWLHDPI
                     EVRVNIEKARARGWVDERAIVEFSKFAAYVAWKFDKYVDFWATFDEPMVTAELGYLAP
```

YVGWPPGILNPKAAKSVIMHQIIAHARAYDSIKKFSDKPVGVILNIIPAYPLNPRDSK
HVKAAENYDLFHNRLFLEAVNKGKLDIDINGEYVKVPHLKRNDWIGNNYYTREVVKYV
EPKYKELPLVTFVGVEGYGYSANPNSISPDNNPTSDFGWEVYPKGLYDSTAEAYEYSE
MVFITENGIADSKDILRPRYIVDHVKEVKRLRENGINVLGYFHWALTDNYEWAMGFKI
RFGLYEVDPITKERIPRRKSVETYKKVVKEGIE"

CDS 1834. .2283
/note="PAB1019"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50448.1"
/db_xref="GI:5458962"
/translation="MPLPPDITFDSVTLLKMHSEKNKRLLEITLAKFTVYLSMLTVQR
YLTARAYLKRNVELEFEVLKDIYNIVPVSEEIILRASKIEANLIRKGVFLDLEDIITA
VTAIETGSLLITDEPKKYEPIRRFGLDTMPLDKFLREVELMVEKEII"

CDS complement(2276. .3079)
/note="PAB1341"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50449.1"
/db_xref="GI:5458963"
/translation="MFENLESFEEAFKKLLEEVLEFDLNDPFKDVEKVLCIEPHPDDC
VIGMGGTIKKLTERGIEVIYACMTDGYMGTLDSSLTGHELATIRRREEEESSKLLGVK
KIYWLNYRDTELPYSREVRKDLVRIIRKEKPDGVFLPDPWLPYEAHPDHRNTGFLALD
AVAFSPLPNFSNVDVEIGLGPHQVSFIALYYTNKPNYFVDITDVMELKLKAIRTHKSQ
FPDDVWEVWEPFLRTVALFYGKKAGVKYAEGFRIMPGLFYHITPFADLI"

CDS 3192. .4340
/note="PAB1020"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50450.1"
/db_xref="GI:5458964"
/translation="MKEVVSSVYKEILVKLGLTEDRIETLEMKGGIIEDEFDGIRYVR
FKDSAGKLRRGTVVIDEEYVIPGFPHIKRIINLRSGIRRIFKRGEFYVEEKVDGYNVR
VVMYKGKMLGITRGGFICPFTTERIPDFVPQEFFKDNPNLILVGEMAGPESPYLVEGP
PYVKEDIQFFLFDVQEIKTGRSLPVEERLKIAEEYGINHVEVFGKYTKDDVDELYQLI
ERLSKEGREGIIMKSPDMKKIVKYVTPYANINDIKIGARVFYELPPGYFTSRISRLAF
YLAEKRIKGEEFERVAKELGSALLQPFVESIFDVEQEEDIHELFKVRVKRIETAYKMV
THFEKLGLKIEIVDIEEIKDGWRITFKRLYPDATNEIRELIGGKAFVD"

CDS 4351. .5658
/note="PAB1021"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50451.1"
/db_xref="GI:5458965"
/translation="MEDIALRLVKAAIRAADPYKAVINSVRVEGNKLLVKDKEFEIKG
KVYVLAFGKAACDMARAIEEILDVEDGVAVTKYGYGKELKKIRVIEAGHPIPDENSIR
GAIEGLKLLEKVKEEDIVFVLISGGGSALFELPEDGISLEDLKLTNELLLKSGAKIHE
INTVRKHISKVKGGKLAKKIKGTGIVLIISDVVGDNLEAIASGPTVKDPTTFQDAKRI
LELYGIWDRVPESVRSHIEKGIRGEAEETLKEELPNIHNFLIASNSISCQAVVEEATK
LGFKAYVMTTTLEGEAKDAGLFIGSIVQEIAKTGNPFEPPVVLVFGGETTVTIEGEGG
KGGPNQELALSATRKIKGLNAILVAFDTDGTDGPTDAAGGIVDGETYEKLRRKGIDIE
KVLKEHNSYEALKKVGSLLFTGPTGTNVNSMIIAIITSKQGRT"

CDS complement(5633. .6655)
/note="PAB1338"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50452.1"
/db_xref="GI:5458966"
/translation="MPIYEARVKVGISSSWVTSKVSWRDALAQLESDRIVVKYLRMGE
VTGEDSFPFSALTDIGIRIPDELKLNPEKDHFGLKFYVPGRGELLVIFTIEENLLIYD
EKKFAEFVHKVFEVLINGKTVMLQLARIVGGAINMESKWEEGWLRVIKVKSARTQRTE
RSVVVITQDKRPVSIFSDLEDIEIEEVEMNGKKVRAWKIRHFHINQSVTSYLYIPDKK
VQLFVLRYLLKYTPSAMEFIIKIADDFPTLKSEFQELMEKELKELESLDEMEKQILVA
LYSGINPLELHQFLGITEKEIEEIYDRMIDKGLLKIVMIRKIVDLTNEGRKIVNKLLK
YGLVSM"

CDS complement(6670. .7425)
/note="PAB1337"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50453.1"
/db_xref="GI:5458967"
/translation="MDVGTLILIAVVVLIGSIVSAWIVSSVISQQMKNMQALMISRSL
EVETKTVEKKRKGTRGSEAKPEEVAKQEVRDVSEEDIVRKIKQIIDEKVHNVANEAKK
KKERLLMLLDVARGYALGFVSEDEYNAFLMRVLSELDEFKRLWLAAYPFPEDKDRLNQ
MLTHVARTKLPIEVREKGKGKEGSVKLSPDEALIRITSSIKSAISILDDLIKKRGDNP
AVTPLEIKLSQELEKCKKKVEDLEKRLQELEAL"

CDS complement(7468. .9687)
/note="PAB1336"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein (tlpC)"
/protein_id="CAB50454.1"
/db_xref="GI:5458968"
/translation="MQFRKKIILMAIMSAVITMAIFGGVTIYSFNKVYGVIKGEIQAP
AIEQGGQLAMSAAEVAEKMFDDFFARVADYGTMATEVVKEAYAKGLDSNETAMREFLL
SEFKKIKELNKDVSYVYFGDEEGHMYMYPPDELPEGYDPRVRPWYVQAKEKGGAIITE
PYRDAATGQWVITYAVPVYVNGQFKGVIGLDVFVSTLLNETKKIKLGKTGYIAIIDQN
GMTLVHPNEEFINKLNIFQVPELKDLADELKKGKDEGYVLYTFQGIKKIGGYKRLKTT
GWIVVATVPLSDITGPILSATDTAINKMKMIVIQGFAVGFVIIVAILVVIYKIASNTL
APLEKLKYAAQALAEGRLKQVSEYLKQIRYLERDEIGALIQAFEAVSKDLVGTLNAIS
KKLERLAEGDLSNGLTVEVRGELRDIIEDIRSVTETFRESIGSLVEMANDLEKRANAL
AQVSKDVTEAINQVNEAIQQVSIEAQRQQETINEITDGMRLVAQTSEESVRAMEEFSG
AVTEVVSIANEGSQKGDEALKRIEDIQHMMSRIEETVSKVAEMSRNIEEITNVITSIA
EQTNLLALNAAIEAARAGEAGRGFAVVAQEIRKLAEESKQAADNIKSIIDKITDEIKE
AVEATKEGVSVIGESSETLRDTIGYLANIATLLQETSERMTTVKEQIVRTQEEVDKAL
RALENLAASAEETTASAEEVSSAIEQQTAAIEELRRAAQELKDMVGRMRQIVGKFKV"

CDS complement(9704. .10195)
/note="PAB1335"
/codon_start=1
/transl_table=11
/product="CHEMOTAXIS PROTEIN CHED"
/protein_id="CAB50455.1"
/db_xref="GI:5458969"
/translation="MTREIKVGIGDYAVGKGEGIISTYGLGSCVGITLYDRVTKVGGL
LHALLPEAARYGHRGNPAKYVDTGLQLLLKEVLKLGASKFRLEAKLFGGAQMFQNIKS
DELKIGERNVQTAKRELKKLGIRLVAEDTGGRGGRTIYLDLSTGKVRMRKVIGGQVIE
KVY"

CDS complement(10179. .10817)
/note="PAB1334"
/codon_start=1
/transl_table=11
/product="CHEA HISTIDINE KINASE"
/protein_id="CAB50456.1"
/db_xref="GI:5458970"
/translation="MKKSELYKDIFKEASNIAMSHALTSLSQMIGGPIEMEAPDVEIV
SRAEFLKLLAERGVSKGFTVMFDVTEGLSGLTILQFPKHSALNISAVLMGMEPGSMEE
LDEMGKSAIMEVGNILISAYTDILSNLIGEPVSLSPPKPAESLYDIEKELGRPDLRNV
NSIVVFKSKFYKEDIGVESYFYLVPTPESFTKLVKKLEKQISEEVEANDEGN"

CDS complement(10814. .11452)
/note="PAB1333"
/codon_start=1
/transl_table=11
/product="CHEA HISTIDINE KINASE related protein"
/protein_id="CAB50457.1"
/db_xref="GI:5458971"
/translation="MDIVAENIEEYVKNLDEFAKSALLETFNIGASHAATALSEMTGR
EVNISVPDLKVVEIKNVPEIVGEDVKVVVYIELGKDFSSHAFFAADYDDVLRMFDVIM
GNPPGTTTELDDMVKSSFMEMGNILISAFANALSQFLGITIEQTPPALTIDFLPAILD
LALADIGKFCDYTIILQTDIKISGVEFEEHFLLFPKPEDMKKILEKLLGGLA"

CDS complement(11436. .13754)
/note="PAB1332"
/codon_start=1
/transl_table=11
/product="chemotaxis histidine kinase (cheA)"
/protein_id="CAB50458.1"
/db_xref="GI:5458972"
/translation="MVTLEDMSQYLDEFLADARDRIDSLSNAILTLEKIVKEGGSEEE
KAELINQIFRDAHTLKGTAATMGFMKLSETAHKMENLFDAIRNSQIELTPEVVDLVLD
FLDAIEAMVNNIEESGSEGDIDVSDLFRRADEFLQGGKGAEKKPEEEKPKEEVEAKEE
EKPEKVEEGVEGKVYHIKVYFQKDAPLKGVRAFLILSDLEERGEVIWTNPGRDVIESG

```
                  NLDKDVIEFKVSSDLPKEELEKVITRHPEVEKVEIIEEGEEEKGPEGEEKEYLIRVYF
                  SPDAPLKGPRSYLILQDLESIGDIVQTNPNRGDIENGNLLEGKFFEVVLRTKEDKSKI
                  VQLIKKHPDVENFVVTEAEAEEKQEKKEEKAEEKKPQPPKQPKKPVIETPRVKISRII
                  KIDVGHLDKLMNLVGELVITKGRLEQIGERLGDRELFEALSTLSRLLTELQDEIMEMR
                  LTPIAEVFNKFPRMVRDLARKMGKEVEFVMEGADIEVDRTILEKLGDALVHLLRNAVD
                  HGIEPPEERVKLGKPRVGRVELIAKREKNHVVIIVRDDGRGIDPEKVKKKAIERGLIT
                  PEEAASLSDEEAINLIFLPGFSTAEKVTDVSGRGVGMDVVKEVVKSMNGSISVQTEVG
                  KGTTFILKLPISMAIIQALLIRVMDETYAIPINNILETIEVDPSILKTIGGKPVIVLR
                  GEIIPVVMLHELFGLPTPQVERFPAIIVDHGAQKVAIGVDELLHKRDIVIKSLGKFLS
                  NVRGFAGATILGDGSVVLIIDIGGLLGGYSG"
     CDS          complement(13717..14823)
                  /note="PAB1331"
                  /codon_start=1
                  /transl_table=11
                  /product="protein-glutamate methylesterase (cheB)"
Query Match            12.1%;  Score 238.2;  DB 2;  Length 265118;
Best Local Similarity  57.1%;  Pred. No. 9e-63;
Matches  495;  Conservative    0;  Mismatches  363;  Indels    9;  Gaps    3;

Qy    1095 aaagccggacagaccgagtgagctttacagtatcccgcttggacaggaagagaaacagct 1154
Db   58841 AGATGCGACAACTCCGACTGAGCTTTACGCTCTCAGGGATGGA---AAAGAGAAGCGTAT 58785
Qy    1155 gactggcgcgaatgacaagtttgtcagggagcatacgatatcaatacctgaagagattca 1214
Db   58784 CACAGACTTTAACGCCTGGATT---AGGGACTATAAACTTTCAAAACCAGAGCACTTCAG 58728
Qy    1215 atatgctacagaagacggcgtgatggtgaacggctggctgatgaggcctgcacaaatgga 1274
Db   58727 GGTTACGGCTAGCGATGGGAAGGAAATAGATGCCTGGATCATGAAGCCCGTTGACTTTAA 58668
Qy    1275 aggtgagacaacatatccacttattcttaacatacacggcggtccgcatatgatgtacgg 1334
Db   58667 GCCTGGAAGGAAGTATCCTGCAGTGCTGGAAATTCACGGTGGGCCAAAAACAGCCTACGG 58608
Qy    1335 acatacatatttcatgagtttcaggtgctggcggcgaaaggatacgcggtcgtttatat 1394
Db   58607 TTATGCATTTATGCACGAGTTTCAGGTTTTAGTTGCCAAAGGCTTCGTTGTAATCTTTTC 58548
Qy    1395 caatccgagaggaagccacggctacgggcaggaatttgtgaatgcggtcagaggagatta 1454
Db   58547 AAATCCAAGGGGAAGTGATGGTTATGGTGAGGAGTTCGCG---GACATAAGGGGTCACTA 58491
Qy    1455 tggggaaaggattatgacgatgtgatgcaggctgtggatgaggctatcaaacgagatcc 1514
Db   58490 CGGTGAGCGCGATTATCAAGACCTAATGGAGGTAGTTGACGAGGCTTTAAAACGCTTTGA 58431
Qy    1515 gcatattgatcctaagcggctcggtgtcacgggcggaagctacggaggttttatgaccaa 1574
Db   58430 CTTTATAGATGAGGAAAGGCTTGGCGTTACCGGAGGATCATACGGAGGCTTCATGACAAA 58371
Qy    1575 ctggatcgtcgggcagacgaaccgctttaaagctgccgttacccagcgctcgatatcaaa 1634
Db   58370 TTGGATCGTTGGGCACACAAACAGATTCAAAGCTGCCGTAACTCAGCGCTCAATATCCAA 58311
Qy    1635 ttggatcagctttcacggcgtcagtgatatcggctatttctttacagactggcagcttga 1694
Db   58310 CTGGATCAGCTTCTTTGGAACGACCGATATCGGCTATTACTTCGCTCCAGATCAGATCGG 58251
Qy    1695 gcatgacatgtttgaggacacagaaaagctctgggaccggtctcctttaaaatacgcagc 1754
Db   58250 TGGCGATCCTTGGAGTAACCTAGAAGGCTACTGGGAGAAGAGCCCATTGAAGTACGCTCC 58191
Qy    1755 aaacgtggagacaccgcttttgatactgcatggcgagcgggatgaccgatgcccgatcga 1814
Db   58190 AAACGTTGAGACTCCGCTCCTGATAATTCACTCCACCGAGGACTACCGCTGTTGGCTTCC 58131
Qy    1815 gcaggcggagcagctgtttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccg 1874
Db   58130 CGAGGCTTTGCAATTTTACATAGCGTTGAAGTATTTGGGCAAAACCGTCGAGCTGGCGAT 58071
Qy    1875 ttttccgaatgcatcgcacaatttatcacgcaccggacacccaagacagcggatcaagcg 1934
Db   58070 ATTCCCAGGTGAGAATCATGACTTGAGTAGGAGTGGAAGGCCAAAGCATAGGGTTAAGAG 58011
Qy    1935 cctgaattatatcagctcatggtttga 1961
Db   58010 GCTCGAGCTCATAGTGGGGTGGTTTGA 57984
```

```
RESULT 4
AX041922/c
LOCUS       AX041922     265118 bp    DNA             PAT       23-NOV-2000
DEFINITION  Sequence 817 from Patent WO0065062.
ACCESSION   AX041922
VERSION     AX041922.1  GI:11340687
KEYWORDS    .
SOURCE      Pyrococcus abyssi.
  ORGANISM  Pyrococcus abyssi
            Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
            Pyrococcus.
REFERENCE   1  (bases 1 to 265118)
  AUTHORS   Forterre,P., Thierry,J.C., Prieur,D., Dietrich,J., Lecompte,O.,
            Querellou,J., Weissenbach,J., Saurin,W., Heilig,R., Flament,D.,
            Raffin,J.P., Henneke,G., Gueguen,Y. and Rolland,J.L.
  TITLE     Genome sequence and polypeptides of pyrococcus _i(abissy), fragment
            and uses thereof
  JOURNAL   Patent: WO 0065062-A 817 02-NOV-2000;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IFREMER
            INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)
FEATURES             Location/Qualifiers
     source          1..265118
                     /organism="Pyrococcus abyssi"
                     /db_xref="taxon:29292"
                     /note="Original length of seq 1: 1765118 (for 812
                     sequences) New sequences : seq 001, from 0.000.001 to
                     0.349.980, length: 349.980 seq 813, from 0.300.001 to
                     0.649.980, length: 349.980 seq 814, from 0.600.001 to
                     0.949.980, length: 349.980 seq 815, from 0.900.001 to
                     1.249.980, length: 349.980 seq 816, from 1.200.001 to
                     1.549.980, length: 349.980 seq 817, from 1.500.001 to
                     1.765.118, length: 265.118"
BASE COUNT    75226 a  61302 c  54005 g  74585 t
ORIGIN

  Query Match            12.1%;  Score 238.2;  DB 9;  Length 265118;
  Best Local Similarity  57.1%;  Pred. No. 9e-63;
  Matches  495;  Conservative    0;  Mismatches  363;  Indels    9;  Gaps    3;

Qy    1095 aaagccggacagaccgagtgagctttacagtatcccgcttggacaggaagagaaacagct 1154
Db   58841 AGATGCGACAACTCCGACTGAGCTTTACGCTCTCAGGGATGGA---AAAGAGAAGCGTAT 58785
Qy    1155 gactggcgcgaatgacaagtttgtcagggagcatacgatatcaatacctgaagagattca 1214
Db   58784 CACAGACTTTAACGCCTGGATT---AGGGACTATAAACTTTCAAAACCAGAGCACTTCAG 58728
Qy    1215 atatgctacagaagacggcgtgatggtgaacggctggctgatgaggcctgcacaaatgga 1274
Db   58727 GGTTACGGCTAGCGATGGGAAGGAAATAGATGCCTGGATCATGAAGCCCGTTGACTTTAA 58668
Qy    1275 aggtgagacaacatatccacttattcttaacatacacggcggtccgcatatgatgtacgg 1334
Db   58667 GCCTGGAAGGAAGTATCCTGCAGTGCTGGAAATTCACGGTGGGCCAAAAACAGCCTACGG 58608
Qy    1335 acatacatatttcatgagtttcaggtgctggcggcgaaaggatacgcggtcgtttatat 1394
Db   58607 TTATGCATTTATGCACGAGTTTCAGGTTTTAGTTGCCAAAGGCTTCGTTGTAATCTTTTC 58548
Qy    1395 caatccgagaggaagccacggctacgggcaggaatttgtgaatgcggtcagaggagatta 1454
Db   58547 AAATCCAAGGGGAAGTGATGGTTATGGTGAGGAGTTCGCG---GACATAAGGGGTCACTA 58491
Qy    1455 tggggaaaggattatgacgatgtgatgcaggctgtggatgaggctatcaaacgagatcc 1514
```

```
Db  58490 CGGTGAGCGCGATTATCAAGACCTAATGGAGGTAGTTGACGAGGCTTTAAAACGCTTTGA 58431

Qy   1515 gcatattgatcctaagcggctcggtgtcacgggcggaagctacggaggttttatgaccaa 1574

Db  58430 CTTTATAGATGAGGAAAGGCTTGGCGTTACCGGAGGATCATACGGAGGCTTCATGACAAA 58371

Qy   1575 ctggatcgtcgggcagacgaaccgctttaaagctgccgttacccagcgctcgatatcaaa 1634

Db  58370 TTGGATCGTTGGGCACACAAACAGATTCAAAGCTGCCGTAACTCAGCGCTCAATATCCAA 58311

Qy   1635 ttggatcagctttcacggcgtcagtgatatcggctatttctttacagactggcagcttga 1694

Db  58310 CTGGATCAGCTTCTTTGGAACGACCGATATCGGCTATTACTTCGCTCCAGATCAGATCGG 58251

Qy   1695 gcatgacatgtttgaggacacagaaaagctctgggaccggtctcctttaaaatacgcagc 1754

Db  58250 TGGCGATCCTTGGAGTAACCTAGAAGGCTACTGGGAGAAGAGCCCATTGAAGTACGCTCC 58191

Qy   1755 aaacgtggagacaccgcttttgatactgcatggcgagcgggatgaccgatgcccgatcga 1814

Db  58190 AAACGTTGAGACTCCGCTCCTGATAATTCACTCCACCGAGGACTACCGCTGTTGGCTTCC 58131

Qy   1815 gcaggcggagcagctgtttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccg 1874

Db  58130 CGAGGCTTTGCAATTTTACATAGCGTTGAAGTATTTGGGCAAAACCGTCGAGCTGGCGAT 58071

Qy   1875 ttttccgaatgcatcgcacaatttatcacgcaccggacacccaagacagcggatcaagcg 1934

Db  58070 ATTCCCAGGTGAGAATCATGACTTGAGTAGGAGTGGAAGGCCAAAGCATAGGGTTAAGAG 58011

Qy   1935 cctgaattatatcagctcatggtttga 1961

Db  58010 GCTCGAGCTCATAGTGGGGTGGTTTGA 57984
```

RESULT 5
E16634
LOCUS E16634 1896 bp DNA PAT 28-JUL-1999
DEFINITION Pyrococcus horikoshi gene for aminoacylase/carboxypeptidase.
ACCESSION E16634
VERSION E16634.1 GI:5711317
KEYWORDS JP 1998210977-A/1.
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1896)
AUTHORS Ishikawa,K., Matsui,I., Ishida,H., Kosugi,Y. and Higuchi,K.
TITLE THERMO-STABLE, ACYLPEPTIDE HYDROLASE, AND GENE CODING THE SAME
JOURNAL Patent: JP 1998210977-A 1 11-AUG-1998;
AGENCY OF IND SCIENCE & TECHNOL
COMMENT OS Pyrococcus horikoshi
PN JP 1998210977-A/1
PD 11-AUG-1998
PF 31-JAN-1997 JP 1997018381
PI ISHIKAWA KAZUHIKO, MATSUI IKUO, ISHIDA HIROYASU, PI KOSUGI YOSHIJI,
PI HIGUCHI KATSUHIKO
PC C12N15/09,C07H21/04,C12N9/78,C12Q1/37,(C12N15/09,C12R1:01), PC (C12N9/78,
PC C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FH
FT source 1. .1896
FT /organism='Pyrococcus horikoshi' FT /strain='JCM9974'
FT mat_peptide 1. .1896
FT /product='aminoacylase/carboxypeptidase'.

FEATURES Location/Qualifiers
source 1. .1896
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 614 a 358 c 491 g 433 t
ORIGIN

Query Match 11.5%; Score 227.2; DB 10; Length 1896;
Best Local Similarity 55.4%; Pred. No. 7.5e-60;
Matches 461; Conservative 0; Mismatches 368; Indels 3; Gaps 1;

```
Qy   1139 aggaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaa 1198

Db   1061 ATGGAAAGGAGAAGAAGGTTACCGACTTTAACAAATGGATAAAGGGTTACACCCTTTCAA 1120

Qy   1199 tacctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatga 1258

Db   1121 AACCTGAACACTTTAAGGTTAAAGCAAGTGACGGGGTTGAAATAGATGCCTGGGTAATGA 1180

Qy   1259 ggcctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtc 1318

Db   1181 AACCGGTGAACTTCAGGAAAGGAAAGAAGTATCCAGCTATTCTAGAGATCCACGGTGGTC 1240

Qy   1319 cgcatatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggat 1378

Db   1241 CTAAAACCGCTTACGGTTACGCTTTTATGCACGAGTTCCACGTTTTAACCTCTAAAGGCT 1300

Qy   1379 acgcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatg 1438

Db   1301 TCGTCGTGATATTCTCAAATCCTAGAGGGAGCGATGGCTACGGAGAGGAGTTCGCGGAT- 1359

Qy   1439 cggtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgagg 1498

Db   1360 --ATAAGGGGACACTATGGGGAGAGGGATTACCAGGATTTAATGGAGGTAGTCGATGAAG 1417

Qy   1499 ctatcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacg 1558

Db   1418 CATTAAGGAGATTTGACTTCATAGATGGGGAAAGGCTAGGAGTTACCGGGGGTTCCTATG 1477

Qy   1559 gaggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttaccc 1618

Db   1478 GTGGCTTCATGACGAACTGGATAGTCGGACATACCAACAGGTTCAAAGCCGCTGTAACCC 1537

Qy   1619 agcgctcgatatcaaattggatcagctttcacggcgtcagtgatatcggctatttcttta 1678

Db   1538 AGAGATCAATTTCAAATTGGATAAGCTTCTTCGGGACAACGGATATAGGTTATTACTTTG 1597

Qy   1679 cagactggcagcttgagcatgacatgtttgaggacacagaaaagctctgggaccggtctc 1738

Db   1598 CTCCAGATCAAATAGGAAAAGATCCCTGGAGCAACTTGGAAGGTTATTGGGAAAAGAGCC 1657

Qy   1739 ctttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatg 1798

Db   1658 CATTAAAGTACGCTCCCAACGTTGAAACTCCCCTGCTTATAATCCACTCTACCGAAGACT 1717

Qy   1799 accgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaagg 1858

Db   1718 ACAGGTGTTGGCTTCCCGAGGCATTGCAACTCTTCATATCCCTAAAATACCTGGGGAAGA 1777

Qy   1859 aaaccaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaa 1918

Db   1778 GAGTTGAATTGGCAATATTCCCAGGAGAAAATCATGACCTAAGTAGATCTGGGAAGCCAA 1837

Qy   1919 gacagcggatcaagcgcctgaattatatcagctcatggtttgatcaacatct 1970

Db   1838 AGCACAGGGTTAAAAGACTTGAACTAATAGCAGGATGGATGGAGAAATGGCT 1889
```

RESULT 6
AP000003
LOCUS AP000003 233000 bp DNA BCT 06-APR-2000
DEFINITION Pyrococcus horikoshii OT3 genomic DNA, 544001-777000 nt. position (3/7).
ACCESSION AP000003 AB009484 AB009485 AB009486 AB009487 AB009488 AB009489
AB009490 AB009491 AB009492 AB009493 AB009494 BA000001

```
VERSION     AP000003.1  GI:3236130
KEYWORDS    .
SOURCE      Pyrococcus horikoshii (strain:OT3) DNA.
  ORGANISM  Pyrococcus horikoshii
            Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
            Pyrococcus.
REFERENCE   1  (sites)
  AUTHORS   Kawarabayasi,Y., Sawada,M., Horikawa,H., Haikawa,Y., Hino,Y.,
            Yamamoto,S., Sekine,M., Baba,S., Kosugi,H., Hosoyama,A., Nagai,Y.,
            Sakai,M., Ogura,K., Otuka,R., Nakazawa,H., Takamiya,M., Ohfuku,Y.,
            Funahashi,T., Tanaka,T., Kudoh,Y., Yamazaki,J., Kushida,N.,
            Oguchi,A., Aoki,K., Nakamura,Y., Robb,T.F., Horikoshi,K.,
            Masuchi,Y., Shizuya,H. and Kikuchi,H.
  TITLE     Complete sequence and gene organization of the genome of a
            hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3
  JOURNAL   DNA Res. 5 (2), 55-76 (1998)
  MEDLINE   98344137
REFERENCE   2  (bases 1 to 233000)
  AUTHORS   Tanaka,T., Kawarabayasi,Y. and Kikuchi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) to the DDBJ/EMBL/GenBank databases. Yutaka
            Kawarabayasi, National Institute of Technology and Evaluation,
            Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo
            151-0066, Japan (E-mail:genomeOT3@nite.go.jp, Tel:+81-3-3481-8951,
            Fax:+81-3-3481-8424)
COMMENT     On Jul 28, 1998 this sequence version replaced gi:3130861
            gi:3130869 gi:3130912 gi:3130938 gi:3130941 gi:3130952 gi:3130999
            gi:3131044 gi:3131081 gi:3131085.
            AB009484-AB009494: submitted (10-DEC-1997)
            Kawarabayasi, Y. is officially affiliated with the National
            Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
            305-0046, Japan.
            Robb, T. F. is at the Center of Marine Biotechnology, University of
            Maryland, Baltimore, MD, USA.
            Horikoshi, K. is at the Japan Marine Science and Technology Center,
            Yokosuka, Kanagawa 237-0061, Japan.
            Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
            Japan.
            Shizuya, H. is at the California Institute of Technology, Pasadena,
            CA, USA.
            The other authors are at the National Institute of Technology and
            Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
            All the sequence with length 100 codons or more between ATG or GTG
            and stop codon are defined as CDS
            Homology analysis is performed by Smith-Waterman algorithm against
            GenBank and GenPept release 103; EMBL release 52.0; SwissProt
            release 34.0; PIR-Protein release 54.0; and OWL release 29.5.
            E-mail address for comments and questions: genomeOT3@nite.go.jp
            Restriction map, ORF organization, sequence alignment and more
            information are available at W.W.W. site of Biotechnology Center,
            URL: http://www.bio.nite.go.jp/.
FEATURES              Location/Qualifiers
     source           1. .233000
                      /organism="Pyrococcus horikoshii"
                      /strain="OT3"
                      /db_xref="taxon:53953"
                      /map="544001-777000 nt. position"
     gene             complement(645. .1067)
                      /gene="PH0605"
     CDS              complement(645. .1067)
                      /gene="PH0605"
                      /note="similar to PIR:A64312 percent identity: 51.111 in
                      136aa; PIR:A64312 percent identity: 51.111 in 136aa;
                      Swiss_Prot:P41375 percent identity: 44.762 in 106aa"
                      /codon_start=1
                      /transl_table=11
                      /product="140aa long hypothetical translation initiation
                      factor eIF-2 beta"
                      /protein_id="BAA29694.1"
                      /db_xref="GI:3257011"
                      /translation="MEIDYYDYEKLLEKAYEELPENVKHHKSRFEVPGALVTIEGNKT
                      IIENFKDIAEALNRDPQHLLKFLLREIATAGTLEGKRVVLQGRFTPYLIANKLKKYIK
                      EYVICPVCGSPDTKIIKRDRFYFLKCEACGAETPIQHL"
     gene             complement(1110. .4436)
                      /gene="PH0606"
     CDS              complement(1110. .4436)
                      /gene="PH0606"
                      /note="similar to Swiss_Prot:P30665 percent
                      identity:41.538 in 331aa. motif=ATP/GTP-binding site motif
                      A (P-loop) ; protein splicing signature"
                      /codon_start=1
                      /transl_table=11
                      /product="1108aa long hypothetical cell division control
                      protein"
                      /protein_id="BAA29695.1"
                      /db_xref="GI:3257012"
                      /translation="MEREEIIERFLKFFREYAEEGEEPLYMGKIKDLLTITPKRSIVI
                      NWMHLNSYDPELAEEVIENPEECILAAEDAIQIILKEDFLREDVPKIHARFHNLPKTL
                      MVKEVGAEHINKLIQVEGVVTRVTEIKPFVSVAVFVCKDCGHEMVVPQKPYEGFVAVK
                      KCEQCGSKNVQLDVEKSKFVNFQMFRIQDRPESLRGGQMPRFIDGILLDDIVDTAMPG
                      DRVIVVGILRVIQEKREKVPTFKKVIEVNHIEPISKELEELEISPEDEQKIRELAKRK
                      DIVSAIVDSIAPAIYGYKEVKKGIALALFGGVSRTLPDGTRLRGDIHVLLVGDPGVAK
                      CVDYDTEVLLGDGRKRKIGEIVEEAIKKAEKEGKLGRVDDGFYAPINLELYALDVRTL
                      KVRKVKADIAWKRTTPEKMLRIRTKRGREIRVTPTHPFFTLEEGRIKTKKAYELKVGE
                      KIATPREEAPEAEIFWDEVVEIEEYKPNNSWVYDLQVPEHHNFIANGIFVHNSQLLRY
                      VANLAPRAIYTSGKSSSAAGLCVAPDTLINTDNGRVEIGKFVEEWMKEVGEISEEGIS
                      YAPCFRKVETFKDGKIVESPIRRVWKLRAPKKLVRIKTENGRSIALTRETKLLTINDG
                      ELSWVEAGEVKVGTYVGTVKSEKDVIPGAGKTIRDVSKLYNMEMEVKDYLTREEVRKA
                      IEKLEEIMNPMNIKIPGVQESYEELLRKLETTNDERVRNETLILLSDVSDAHELAKEK
                      IEKIKEIVNSEVHWEKVTEVGEVDGVEYVYDLTVEGSHNFVANGFIVHNTAAVVRDEF
                      TGGWVLEAGALVLADGGYALIDELDKMSDRDRSVIHEALEQQTISLSKAGITATLNAR
                      TTVIAAANPKHGRFNRMKRVSEQIDLPPTLLSRFDLIFVLMDEPDEKVDSEIARHILK
                      VRRGESEVVTPKIPHELLRKYIAYARKNIHPVISEEAMEEIEKYYVKMRRSAKKSSGD
                      EIKPIPITARQLEALIRLSEAHARMRLSPIVTREDAREAIKLMEYTLKQIAVDEAGQL
                      DVTILEVGQSARKLSKVERILDIVEKLERTSEKGAKIDDILEEAAKFGIEKSEAREIL
                      EKLLEKGQIYMPENGYYKTV"
     gene             1630. .1992
                      /gene="PH0607"
     CDS              1630. .1992
                      /gene="PH0607"
                      /note="motif=glutamine amidotransferases class-II active
                      site"
                      /codon_start=1
                      /transl_table=11
                      /product="120aa long hypothetical protein"
                      /protein_id="BAA29696.1"
                      /db_xref="GI:3257013"
                      /translation="MYFLSNSCGIFGVTTSLSPLLTFNICLAISLSTFSSGSSMRTKI
                      KSNLLSNVGGRSICSDTLFILLNLPCFGFAAAITVVLAFSVAVMPAFDRLIVCCSRAS
                      WITLLSLSLILSSSSINA"
     gene             3592. .3981
                      /gene="PH0608"
     CDS              3592. .3981
                      /gene="PH0608"
                      /codon_start=1
                      /transl_table=11
                      /product="129aa long hypothetical protein"
                      /protein_id="BAA29697.1"
                      /db_xref="GI:3257014"
                      /translation="MAETISFLFASSLIFCSSSGDISSSSSSLLMGSMWLTSITFLNV
                      GTFSLFSWITLRIPTTITLSPGIAVSTISSRSIPSMNLGICPPLKLSGLSCILNIWKF
                      TNLLFSTSSWTFLLPHCSHFFTATKPS"
     gene             3693. .4040
                      /gene="PH0609"
     CDS              3693. .4040
                      /gene="PH0609"
                      /codon_start=1
                      /transl_table=11
                      /product="115aa long hypothetical protein"
                      /protein_id="BAA29698.1"
                      /db_xref="GI:3257015"
                      /translation="MVDLYHLLERWNFLSLFLDHSEDPHDYYPITWHCCIHYIIQEYP
                      IYEPRHLPSSQALWPILYSEHLEVHEFALLHIKLDVFTSALLAFLYCNKALIGLLRHN
                      HFMPAILTNEDRD"
     gene             complement(4439. .4936)
                      /gene="PH0610"
```

```
CDS             complement(4439. .4936)
                /gene="PH0610"
                /note="motif=DEAD and DEAH box families ATP-dependent
                helicases signatures"
                /codon_start=1
                /transl_table=11
                /product="165aa long hypothetical protein"
                /protein_id="BAA29699.1"
                /db_xref="GI:3257016"
                /translation="MEMLTGKALIPVKVLRSFENWKEGDMILIEDWKARELWELGIVE
                IIDEADKVIGEIDKVLEEERKNAPISAIPENLYERAEFYIYYLEKYIQKGGEGNIDVI
                HTKLTKLKNLKKKYKMLKDIRFKKILEAVRLRPNSMEILSRLAPQERRIYLQISRIRN
                EWIGE"
gene            complement(4979. .5737)
                /gene="PH0611"
CDS             complement(4979. .5737)
                /gene="PH0611"
                /note="similar to Swiss_Prot:P46577 percent
                identity:30.736 in 238aa. motif=ATP/GTP-binding site motif
                A (P-loop)"
                /codon_start=1
                /transl_table=11
                /product="252aa long hypothetical protein"
                /protein_id="BAA29700.1"
                /db_xref="GI:3257017"
                /translation="MGGTMIIVFVGTAGSGKTSLTGAFGEYLEENYKVAYVNLDTGVK
                ELPYKPDIDVREFVTVEEIMKEGYGPNGAIVESYDRLMNKFDYYLKEILELESKKDYV
                LIDTPGQMETFLFHEFGVKLMENLPYPLVVYLSDPEILRKPTDYCFVRFFALLIDLRL
                GATTVPALNKVDLLKEEQLEKHRKYFEDLDYLMGRLKFDPSMQGLMAYKMCSIMMEVL
                PPIRVLYLSAKTREGFEDLETLAYEHYCTCGDLT"
gene            complement(5768. .6505)
                /gene="PH0612"
CDS             complement(5768. .6505)
                /gene="PH0612"
                /note="similar to PIR:C64368 percent identity: 41.379 in
                233aa; PIR:B64321 percent identity: 35.841 in 227aa;
                owl:S76970 percent identity: 34.188 in 239aa.
                motif=ATP/GTP-binding site motif A (P-loop)"
                /codon_start=1
                /transl_table=11
                /product="245aa long hypothetical cell division inhibitor
                MinD"
                /protein_id="BAA29701.1"
                /db_xref="GI:3257018"
                /translation="MTRIISIVSGKGGTGKTTVTANLSVALGEMGRKVLAVDGDLTMA
                NLSLVLGVDDVNITLHDVLAGDAKLEDAIYMTQFENVYILPGAVDWEHVIKADPRKLP
                EVIKSLKGKYDFILIDCPAGLQLDAMSAMLSGEEAILVTNPEISCLTDTMKVGMVLKK
                AGLAILGFILNRYGRSERDIPPEAAQDVMDVPLLAVIPEDPVIREGTLEGIPAVKYKP
                ESKGAQAFIKLAEEVDKLAGIKAKIMY"
```

```
Query Match             11.5%;  Score 227.2;  DB 2;  Length 233000;
Best Local Similarity   55.4%;  Pred. No. 2.6e-59;
Matches  461;  Conservative    0;  Mismatches  368;  Indels    3;  Gaps    1;

Qy   1139 aggaagagaaacagctgactggcgcgaatgacaagtttgtcagggagcatacgatatcaa 1198
          | | | || |  ||  |  |  ||    | |||| |   | | ||   | ||  | ||||
Db 228212 ATGGAAAGGAGAAGAAGGTTACCGACTTTAACAAATGGATAAAGGGTTACACCCTTTCAA 228271

Qy   1199 tacctgaagagattcaatatgctacagaagacggcgtgatggtgaacggctggctgatga 1258
          |||||||  |  || |   |    ||  |||||  ||      |  | | |||| | ||||
Db 228272 AACCTGAACACTTTAAGGTTAAAGCAAGTGACGGGGTTGAAATAGATGCCTGGGTAATGA 228331

Qy   1259 ggcctgcacaaatggaaggtgagacaacatatccacttattcttaacatacacggcggtc 1318
          || |   |  |       |  |  |  ||||||  ||||||  |  || ||||| ||||
Db 228332 AACCGGTGAACTTCAGGAAAGGAAAGAAGTATCCAGCTATTCTAGAGATCCACGGTGGTC 228391

Qy   1319 cgcatatgatgtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggat 1378
          |  | |     |||||  |  | | | | || ||||| || ||  |  |  | |||||  |
Db 228392 CTAAAACCGCTTACGGTTACGCTTTTATGCACGAGTTCCACGTTTTAACCTCTAAAGGCT 228451

Qy   1379 acgcggtcgtttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatg 1438
            ||  ||  | |     ||||| ||||| ||| | ||||||||  |||| || | | ||
Db 228452 TCGTCGTGATATTCTCAAATCCTAGAGGGAGCGATGGCTACGGAGAGGAGTTCGCGGAT- 228510

Qy   1439 cggtcagaggagattatgggggaaaggattatgacgatgtgatgcaggctgtggatgagg 1498
             | || ||| | |||||||  | ||||||  | ||| | ||| |||  || |||||  |
Db 228511 --ATAAGGGGACACTATGGGGAGAGGGATTACCAGGATTTAATGGAGGTAGTCGATGAAG 228568

Qy   1499 ctatcaaacgagatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacg 1558
          |  | |   ||  |      || |||     |  |||| || || || || ||   ||| |
Db 228569 CATTAAGGAGATTTGACTTCATAGATGGGGAAAGGCTAGGAGTTACCGGGGGTTCCTATG 228628

Qy   1559 gaggttttatgaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttaccc 1618
          | || || ||||| |||||||| ||||| || || ||| | || ||||| || || ||||
Db 228629 GTGGCTTCATGACGAACTGGATAGTCGGACATACCAACAGGTTCAAAGCCGCTGTAACCC 228688

Qy   1619 agcgctcgatatcaaattggatcagctttcacggcgtcagtgatatcggctatttcttta 1678
          || | || || |||||||||||| |||||   |||    |  ||||| || |||| ||||
Db 228689 AGAGATCAATTTCAAATTGGATAAGCTTCTTCGGGACAACGGATATAGGTTATTACTTTG 228748

Qy   1679 cagactggcagcttgagcatgacatgtttgaggacacagaaaagctctgggaccggtctc 1738
          |     ||  | |   | ||    |       ||   |||      |||||   |  |
Db 228749 CTCCAGATCAAATAGGAAAAGATCCCTGGAGCAACTTGGAAGGTTATTGGGAAAAGAGCC 228808

Qy   1739 ctttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatg 1798
          | ||||| |||||  | |||||  || || || ||  | ||| | ||          ||
Db 228809 CATTAAAGTACGCTCCCAACGTTGAAACTCCCCTGCTTATAATCCACTCTACCGAAGACT 228868

Qy   1799 accgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaagg 1858
          || | ||   | |     ||||  ||| || || ||  | || ||| |  |||| |||
Db 228869 ACAGGTGTTGGCTTCCCGAGGCATTGCAACTCTTCATATCCCTAAAATACCTGGGGAAGA 228928

Qy   1859 aaaccaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaa 1918
            |   | | |     || ||    | |   ||  |  ||     |  | ||  | ||||
Db 228929 GAGTTGAATTGGCAATATTCCCAGGAGAAAATCATGACCTAAGTAGATCTGGGAAGCCAA 228988

Qy   1919 gacagcggatcaagcgcctgaattatatcagctcatggtttgatcaacatct 1970
            ||  || | ||  | ||  |    ||      |||| | ||  ||   ||
Db 228989 AGCACAGGGTTAAAAGACTTGAACTAATAGCAGGATGGATGGAGAAATGGCT 229040
```

```
RESULT   7
AE001879/c
LOCUS       AE001879    10256 bp    DNA             BCT       22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 16 of 229 of the complete
            chromosome 1.
ACCESSION   AE001879 AE000513
VERSION     AE001879.1  GI:6457832
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE   1  (bases 1 to 10256)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et,al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 10256)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10256
                     /organism="Deinococcus radiodurans"
```

/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
gene 100. .1017
/gene="DR0164"
CDS 100. .1017
/gene="DR0164"
/note="similar to GB:D14846 PID:286165 percent identity: 49.11; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endo alpha-1,4 polygalactosaminidase, putative"
/protein_id="AAF09753.1"
/db_xref="GI:6457833"
/translation="MRLPLTVLLPLSAVLLASCGGTPSPVAATVFPATPAAQPTRTLA
AQATPPLRLPPAGKLAWDWQIGAATEAKVALPAGVSLLDLDGFETSAAKVADLKAQGV
YTVCYLNVGSYESYRPDAAQYPDSLKIQTDPNWPDESFVDIRDVFREGGVLAGILDRR
LALCAAKGFDAVEPDNLQNDQNVTSGVISRQDQLDFNGWLADRAHAHGLAILQKNGPD
YVLQADRQGRLMVDLFDGVLNESCQRYKECGPLTEYVRRGKLALNVEYRQADLNCATM
NSLGVNALFKDLSLVGKLQKAYKRVSCTG"
gene complement(1079. .3046)
/gene="DR0165"
CDS complement(1079. .3046)
/gene="DR0165"
/note="similar to GB:AL009126 percent identity: 56.01; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl-peptide hydrolase, putative"
/protein_id="AAF09754.1"
/db_xref="GI:6457834"
/translation="MNNSETPAPGPDSLLALAFPSDPQVSPDGKQVAFVLAQISEEDP
AKPDKDFARPRYRSGLWLSEGGAARPLTHAETGRGDSAPRWSPDGQNLAFVRSAGEVK
AALMLLPLKGGEARRVTHFKNGVSGPQWSPDGRFIAFTTTADTEDKRDERGEARVLTR
PVYRANGADWLPERPAALWLYDVEADKLREWYAPEIGIGALSWWPDSRGVLIVQSEDE
WQASQWRQDVYDLPLPTADAPAAPQKLLDWNSAAHGLAPHPDGQRFALIGRPAGKGNT
EHAHLYLIENGQHRRLDTGHDHPVGDAVGGDCHVGAFPEGPRWLDGDTLLFSSTVRGS
VGLFTAHIGGGVKAYDHDPQGVISAFTANEHGVALIRESATRFPEVELNGQRVTDLHA
RFPFPVREPQRVTFETELGEGEGWVLLPEGEQKVPALLNIHGGPHTDYGHGFTHEFQL
MAARGYGVCYSNPRGSVGYGQAWVDAIYGRWGTVDADDLLNFFDRCLEAVPRLDAAKT
AVMGGSYGGFMTNWITGHTTRFQAAITDRCISNLISFGGTSDIGLRFWDDELGLDFSR
RADALKLWDLSPLQYVENVKTPTLIVHSVLDHRCPVEQAEQWYAALHKHQVPVRFVRF
PEENHELSRSGRPDRRLTRLNEYFAWLERWL"
gene 3098. .3352
/gene="DR0166"
CDS 3098. .3352
/gene="DR0166"
/note="similar to SP:P31787 PID:1323021 PID:1620454 PID:170970 percent identity: 66.09; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl-CoA-binding protein"
/protein_id="AAF09755.1"
/db_xref="GI:6457835"
/translation="MNFEQAQQEVNTLSKKPGNDVLLKLYALYKQGSAGDVDGKRPGG
FDFVGGAKYDAWAGLKGKTQEQAQQEYVALVEELKAKDGQ"
gene 3663. .4649
/gene="DR0167"
CDS 3663. .4649
/gene="DR0167"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF09762.1"
/db_xref="GI:6457842"
/translation="MPSANVSPPCPSGVRGGGMGPKAKAEASKPHPQIPVKLPFVTAP
DALAAAKARMRDLAAAYVAALPGRDTHSLMAGVPGVDLKFMPLGWRDGAFDPEHNVIL
INSAARPERQRFTLAHEIGHAILLGDDDLLSDIHDAYEGERLEQVIETLCNVAAAAIL
MPEPVIAEMLERFGPTGRALAELAKRAEVSASSALYALTEQTPVPVIYAVCAPGKPPR
EQAASDEDAGPSTEKVLTVRASSSTRGVKYTLASGTPVPADHPAALALATGMEVREES
YVPFRSGRKMKAEVDAYPSRGIVAVSFEFDPARLGRKDSEQADRDEPQDAAQ"

gene 4646. .5515
/gene="DR0168"
CDS 4646. .5515
/gene="DR0168"
/note="similar to GB:L06494 SP:P26282 GB:U01376 GB:X68776 GB:X68777 percent identity: 62.16; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydropteroate synthase"
/protein_id="AAF09756.1"
/db_xref="GI:6457836"
/translation="MNWTGRIHELTFGFPVPSAEKGPNGWTVRWPGCRVMGILNATPD
SFSDGGQHLQLDAALATARRMRDTGVFILDIGGESTRPGAEPVDAATELDRVLPLIRA
LRGEHVLLSVDTMKPEVAAEALRAGAHLVNDVTGLRDPAMIRVCADAGAAACVMHMQG
EPRTMQHQPVYSDVVAEVHGFLRERAAEVKSAGVPSVLLDPGIGFGKTLDHNLSLLRA
VDDLASGPDPVLIAASRKKTIDLIAGVPEAADRDPGSLALHLHAARCGAALVRLHAAG
EMVQALRVEAALA"
gene 5540. .5896
/gene="DR0169"
CDS 5540. .5896
/gene="DR0169"
/note="similar to GB:D26185 SP:P28823 GB:M34053 PID:143411 PID:467467 percent identity: 62.18; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydroneopterin aldolase"
/protein_id="AAF09757.1"
/db_xref="GI:6457837"
/translation="MSRVVLEGLEFHAHHGVFAAEGVLGARFVVDAELHSPFAGIADD
LGQAVNYAAAYAAIREEVTEKTHQLIEVLADRIADRLLSDFSRLERVKVRVHKPFAPL
PGIFRDVYAEIEKERT"
gene 5893. .6408
/gene="DR0170"
CDS 5893. .6408
/gene="DR0170"
/note="similar to GB:D26562 SP:P26281 GB:L06495 PID:146013 PID:473801 percent identity: 56.86; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase"
/protein_id="AAF09758.1"
/db_xref="GI:6457838"
/translation="MTRAFIALGANLGDPQATLRRALTELGALGEVRGVSALYRTAPV
GGPPQGPDYLNAVAELETALSAPALLAALHALEADAGRTRDVRWEARVLDLDLIVYGE
DVSDDPALTLPHPRAWERGFVLAPLSDLAPSLAHPRTGEMVAQALRRVDLAGVEVSEL
SLCWPELGSGE"
gene complement(6362. .7054)
/gene="DR0171"
CDS complement(6362. .7054)
/gene="DR0171"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF09763.1"
/db_xref="GI:6457843"
/translation="MNLEGVHPLRLGLCWCEMEQTSRSSSQQVTDTEVASALLDTRSR
LVLSAFLTPTSLSAAARDLGRPLNTVKYQWQKLRRLGLLKMVCDGRPPLYQAEGESFF
IPYELTPAHWPQDMLQLLHTEWNRRMDTALLSTAEDEARRHGVWGMQVRRQENALTLE
HAAWSAGWNFTRPQAPAVLDLLGELSLDFEQAKALQRDLLQLLAQYRRKPGSQPHLIR
LTLTPASTEKAH"
gene 7034. .7849
/gene="DR0172"
CDS 7034. .7849
/gene="DR0172"
/note="similar to SP:Q50665 PID:1449319 GB:AL123456 percent identity: 49.77; identified by sequence similarity; putative"
/codon_start=1

```
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF09759.1"
                /db_xref="GI:6457839"
                /translation="MDTLKIHGFSTGQIQVKSKFLRPEPQNLLGSLRSFLDRQYSAPL
                PVWCWLIEHPSGLIVVDTGLTCTARSNWFSRTQTRLHYTEHDRLSRQMQRAGFDPAQV
                RVLILTHLHIDHDRGMGDFPEAHLLVSARELRDYRSPLARLMGYSAPTWKSVPHAAPF
                LPLAHGPFRESADVMGDGRVLLLPTPGHTPGHLSVLVRQPGHDILLAGDLTYSEQDLR
                TGRLGGIVPSPHIHRKSMRQVRQLAQLRPLIYLPSHDPDSLRRLQRNQPFSLT"
   gene         complement(7904. .8999)
                /gene="DR0173"
                /note="This region contains an authentic frame shift and
                is not the result of a sequencing artifact; similar to
                SP:P28367 PID:216335 PID:2331287 GB:AL009126 percent
                identity: 66.01; identified by sequence similarity;
                putative"
   gene         9254. .9778
                /gene="DR0174"
   CDS          9254. .9778
                /gene="DR0174"
                /note="similar to SP:P02410 PID:42855 PID:606170 GB:U00096

Query Match              9.7%;  Score 191.2;  DB 1;  Length 10256;
Best Local Similarity   55.7%;  Pred. No. 2.6e-48;
Matches  389;  Conservative    0;  Mismatches  303;  Indels    6;  Gaps    1;

Qy   1272 ggaaggtgagacaacatatccacttattcttaacatacacggcggtccgcatatgatgta 1331
Db   1787 GGAAGGTGAGCAGAAAGTGCCCGCGCTGCTCAATATTCACGGCGGGCCGCACACCGACTA 1728

Qy   1332 cggacatacatattttcatgagtttcaggtgctggcggcgaaaggatacgcggtcgttta 1391
Db   1727 CGGGCACGGCTTTACCCACGAGTTCCAGCTGATGGCGGCGCGCGGCTACGGCGTGTGCTA 1668

Qy   1392 tatcaatccgagaggaagccacggctacgggcaggaatttgtgaatgcggtcagaggaga 1451
Db   1667 CTCCAACCCGCGCGGCTCGGTGGGCTACGGGCAGGCGTGGGTGGACGCCATCTACGGGCG 1608

Qy   1452 ttatgggggaaaggattatgacgatgtgatgcaggctgtggatgaggctatcaaacgaga 1511
Db   1607 CTGGGGCACGGTGGACGCGGACGACCTGCTGAACTTCTTCGACCGCTGCCTGGAAGCGGT 1548

Qy   1512 tccgcatattgatcctaagcggctcggtgtcacgggcggaagctacggaggttttatgac 1571
Db   1547 GCCGCGCCTGGACGCCGCAAAAACCGCCGTCATGGGCGGCAGCTACGGCGGCTTCATGAC 1488

Qy   1572 caactggatcgtcgggcagacgaaccgctttaaagctgccgttacccagcgctcgatatc 1631
Db   1487 CAACTGGATCACCGGGCACACCACCCGTTTTCAGGCCGCCATCACCGACCGCTGCATCAG 1428

Qy   1632 aaattggatcagctttcacggcgtcagtgatatcggctatttctttacagactggcagct 1691
Db   1427 CAACCTGATTTCGTTCGGGGGCACGTCCGACATCGGGCTGCGCTTCTGGGACGACGAGCT 1368

Qy   1692 tgagcatgacatgt------ttgaggacacagaaaagctctgggaccggtctcctttaaa 1745
Db   1367 GGGGCTGGACTTCTCGCGCCGGGCGGACGCGCTGAAGCTCTGGGACCTTTCGCCACTGCA 1308

Qy   1746 atacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatgaccgatg 1805
Db   1307 ATACGTGGAAAACGTGAAGACGCCGACCCTGATCGTCCATTCGGTCCTTGACCACCGCTG 1248

Qy   1806 cccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaaggaaaccaa 1865
Db   1247 CCCTGTCGAGCAAGCCGAGCAGTGGTACGCCGCGCTGCACAAGCATCAGGTGCCGGTGCG 1188

Qy   1866 gcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaagacagcg 1925
Db   1187 CTTCGTGCGCTTCCCCGAGGAAAACCACGAACTGTCGCGCTCGGGCCGGCCCGACAGACG 1128

Qy   1926 gatcaagcgcctgaattatatcagctcatggtttgatc 1963
Db   1127 GCTGACGCGGCTGAACGAGTACTTCGCCTGGCTGGAAC 1090
```

```
RESULT   8
AP000062
LOCUS       AP000062   251700 bp    DNA             BCT       06-APR-2000
DEFINITION  Aeropyrum pernix genomic DNA, section 5/7.
ACCESSION   AP000062 BA000002
VERSION     AP000062.1  GI:5105244
KEYWORDS    .
SOURCE      Aeropyrum pernix (strain:K1) DNA.
  ORGANISM  Aeropyrum pernix
            Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
            Aeropyrum.
REFERENCE   1  (sites)
  AUTHORS   Kawarabayasi,Y., Hino,Y., Horikawa,H., Yamazaki,S., Haikawa,Y.,
            Jin-no,K., Takahashi,M., Sekine,M., Baba,S., Ankai,A., Kosugi,H.,
            Hosoyama,A., Fukui,S., Nagai,Y., Nishijima,K., Nakazawa,H.,
            Takamiya,M., Masuda,S., Funahashi,T., Tanaka,T., Kudoh,Y.,
            Yamazaki,J., Kushida,N., Oguchi,A., Aoki,K., Kubota,K.,
            Nakamura,Y., Nomura,N., Sako,Y. and Kikuchi,H.
  TITLE     Complete genome sequence of an aerobic hyper-thermophilic
            crenarchaeon, Aeropyrum pernix K1
  JOURNAL   DNA Res. 6 (2), 83-101 (1999)
  MEDLINE   99310339
REFERENCE   2  (bases 1 to 251700)
  AUTHORS   Tanaka,T., Hino,Y., Kawarabayasi,Y. and Kikuchi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1998) to the DDBJ/EMBL/GenBank databases. Yutaka
            Kawarabayasi, National Institutte of Technology and Evaluation,
            Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo
            151-0066, Japan (E-mail:kyutaka@kazusa.or.jp, Tel:+81-3-3481-8951,
            Fax:+81-3-3481-8424)
COMMENT     Kawarabayasi, Y. is officially affiliated with the National
            Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
            305-0046, Japan.
            Nomura, N. and Sako, Y. are at Kyoto University, Sakyo-ku, Kyoto
            606-8502, Japan
            The other authors are at the National Institute of Technology and
            Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
            All the sequence with length 100 codons or more between ATG or GTG
            and stop codon are defined as CDS
            Homology analysis is performed by Smith-Waterman algorithm against
            GenBank and GenPept release 109; EMBL release 56.0; SwissProt
            release 36.0; PIR-Protein release 57.0; and OWL release 31.0.
            E-mail address for comments and questions: genomeAPE@nite.go.jp
            Restriction map, ORF organization, sequence alignment and more
            information are available at W.W.W. site of Biotechnology Center,
            URL: http://www.mild.nite.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .251700
                     /organism="Aeropyrum pernix"
                     /strain="K1"
                     /db_xref="taxon:56636"
     tRNA            complement(948. .1026)
                     /gene="APEtRNA28"
                     /note="codon recognized: GGG; anticodon:CCC"
                     /product="tRNA-Gly"
                     /anticodon=(pos:988. .990,aa:Gly)
     gene            complement(948. .1026)
                     /gene="APEtRNA28"
     gene            1157. .2149
                     /gene="APE1559"
     CDS             1157. .2149
                     /gene="APE1559"
                     /note="similar to PIR:A69437 percent identity:31.228 in
                     285aa."
                     /codon_start=1
                     /transl_table=11
                     /product="330aa long hypothetical protein"
                     /protein_id="BAA80558.1"
                     /db_xref="GI:5105245"
                     /translation="MSGPGTYSPPSSRLQYLGCCAGVAGARLVPGPLGGRFPFSNCVL
                     VEDAGFKVLVDTGCEPWERGLVDVDVVVFTHFHPDHIRGFHRLRAGRVAAPEGEEPYA
```

GRLERLALRFAGERWREWISMARLVMDLEGVPEPGEYFRPGEDLCFRGVCIKTYPARG
HLQTHTLLELPGGLVHLVDIDLTGFGPWYGNPEASPELFLLDIEAAAEMEARGYLSAH
KDRVYSRGEGVEALARFAGRLVEQAEGVYKALEAAGRPLWPWELTGRGIIYRRYLEGM
EAIMSYFEGVMIEKLLGLLGLWGCTRKTPGGYVARDCHLGVIKDRITRNILSLA"

gene complement(1196. .1684)
/gene="APE1560"

CDS complement(1196. .1684)
/gene="APE1560"
/codon_start=1
/transl_table=11
/product="162aa long hypothetical protein"
/protein_id="BAA80559.1"
/db_xref="GI:5105246"
/translation="MDKPPGQLEQGVGLEVPPSGVGLNADPPEAEVLPRPEILPRLRH
AFQIHNQPSHADPLPPPLAGKPEGKPLKPPSIGLLTLRGSHPASPQPMEPPDMVWVEV
SEHHYINIHQPTLPGLTPRIHENLEPRILNQHTVAKGKPPPKRPRNKPSTSHPSTAPQ
IL"

gene 2258. .2437
/gene="APES056"

CDS 2258. .2437
/gene="APES056"
/note="similar to OWL:SOU61050 percent identity:41.667 in
24aa."
/codon_start=1
/transl_table=11
/product="59aa long hypothetical protein"
/protein_id="BAA80560.1"
/db_xref="GI:5105247"
/translation="MARVRIVAKKNGPYLLEVDGKVQTALCRCGHSNNKPYCDGTHAR
VGFQADEKVAFEAEF"

gene 2536. .2952
/gene="APE1561"

CDS 2536. .2952
/gene="APE1561"
/note="motif=neutral zinc metallopeptidases, zinc-binding
region signature"
/codon_start=1
/transl_table=11
/product="138aa long hypothetical protein"
/protein_id="BAA80561.1"
/db_xref="GI:5105248"
/translation="MRYERAPDVCRAAEVLVDALGLDYIDVSRVYCVRSWGSRGRAYA
RIYGTPGAWAAALGYNPGYVIEVVSEKFDPLPLEEKVKVVIHELLHIPRTFSGGLRAH
GPLVNGRRVAALYRRVRGDARVMGILREALDGGPQE"

gene complement(2599. .3063)
/gene="APE1562"

CDS complement(2599. .3063)
/gene="APE1562"
/codon_start=1
/transl_table=11
/product="154aa long hypothetical protein"
/protein_id="BAA80562.1"
/db_xref="GI:5105249"
/translation="MQGSRVSWPGRAARPGYAGGLGSWPFSPWPGPRPCWELFLRASI
QGFAEDAHNPGITPDPPVQRCNPPPVNQGAMGPEAAAEGAGYVEQLMYDDLNLLLQWQ
RVELLRDNLYNVAGVIAEGRSPSPRRPVDPRISPPPTPPAPHAVHPGDVDIV"

gene complement(2906. .3376)
/gene="APE1563"

CDS complement(2906. .3376)
/gene="APE1563"
/codon_start=1
/transl_table=11
/product="156aa long hypothetical protein"
/protein_id="BAA80563.1"
/db_xref="GI:5105250"
/translation="MEKEDGVLRLRTTPPKGCEDLTPDVHAYLGSLGVREIALGYTRV
YCRDGAAEVLLDLEGMLTLAQPEPGEIMAATVTLSPGNAILLITLQERQDGATAASTA
KGDAGEPGVVAREGREAGLRWGLGKLAILAMAGAAALLGAILEGLHPRLRGGCP"

gene complement(3615. .4268)
/gene="APE1564"

CDS complement(3615. .4268)
/gene="APE1564"
/note="similar to PIR:H70375 percent identity:45.455 in
110aa. motif=C-5 cytosine-specific DNA methylases
signatures"
/codon_start=1
/transl_table=11
/product="217aa long hypothetical protein"
/protein_id="BAA80564.1"
/db_xref="GI:5105251"
/translation="MAGVGGRPPPSVGSVVVADTTMFTDAKLREVLGARSLEEAARIM
ARILARASRMGLTVYITPSVRAEIERFMLGNGVHPTTIARLMAWVRVKPPTTHELRLP
AAMFRRYVETVRARLDKGLRVAEDHVRRALRSRDEEGDIVRSLREKYREYTRKGMLDS
VEDVDTLLLALELGAAIVTSDEGLRRAAEDLGLTVLTPVELLEYILALEEEVKSVEE"

gene 3888. .4304
/gene="APE1565"

CDS 3888. .4304
/gene="APE1565"
/codon_start=1
/transl_table=11
/product="138aa long hypothetical protein"
/protein_id="BAA80565.1"
/db_xref="GI:5105252"
/translation="MILSHPKPLVEPGPNGLHIPSEHGCWEAQLMRGGRLNPHPRHKP
GNGCRVHAVPQHEPLYLSPDARSYVHRQPHPGSPGQYPSHYPGSLLQAPGPEDLAQLG
VGKHGRVCHHNAAHARRRPPAHASHPLQPPCPSLLV"

gene 3955. .4299
/gene="APE1566"

CDS 3955. .4299
/gene="APE1566"
/note="motif=prokaryotic membrane lipoprotein lipid
attachment site"
/codon_start=1
/transl_table=11
/product="114aa long hypothetical protein"
/protein_id="BAA80566.1"
/db_xref="GI:5105253"
/translation="MAAGRRSSCVVGGLTLTHAISLAMVVGCTPFPSMNLSISALTLG
VMYTVSPILEARASILAIILAASSRLLAPRTSRSLASVNMVVSATTTLPTLGGGLPPT
PATPSNPHVPHF"

gene complement(4296. .5405)
/gene="APE1567"

CDS complement(4296. .5405)
/gene="APE1567"
/note="similar to Swiss_Prot:Q57857 percent
identity:33.430 in 344aa. motif=aminoacyl-transfer RNA
synthetases class-II signatures"
/codon_start=1
/transl_table=11
/product="369aa long hypothetical protein"
/protein_id="BAA80567.1"
/db_xref="GI:5105254"
/translation="MDPGSVDLEALSTRRSVRVSRFEDVVYVGFRRQFRGVPEGTLVA
FRRGEQIVVWGYPSIKRMLLPRVAVPRWFPGPTVLVEEKMNGYNVRVFTLGGMVYAAT
RGGLICPYTTRRLRRLYGGALQKILEDLGAEGSFIAGEVVGLENPYTRYYYEEAPGFG
YFIFDIFKGGRQLPPRVKFSLAPEYGLKTVNLLAAEIPATASGVERLYTIVEDLEKRGR
EGVIVKDPEGRVEPLKYTTSRTNIGDIRLGMRYPFEEGRSFLFPRILREIFREWETGR
RRYGELGEAILAPAIEAVEAVSRGGRLVEEFELVFANEVEAEEVIAYFASLGVHLEIA
GVARGVDGVRVAFRKPRKSEGEIARILETGISPLD"

gene complement(5525. .5854)
/gene="APE1568"

Query Match 8.2%; Score 161.8; DB 2; Length 251700;
Best Local Similarity 54.5%; Pred. No. 1.1e-38;
Matches 369; Conservative 0; Mismatches 302; Indels 6; Gaps 2;

Qy 1297 attcttaacatacacggcggtccgcatatgatgtacggacatacatattttcatgagttt 1356
Db 172547 ATTCTCTATATACATGGGGGGCCTAAGACGAGCTACGGCTACGCCTTCATACACGAGTTC 172606

Qy 1357 caggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagccacggc 1416
Db 172607 CAAATGCTAGCCTCTCAGGGCTTTGCGGTGATTTACTCGAACCCCAGGGGGAGTGACGGT 172666

Qy 1417 tacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgat 1476

```
Db 172667 TACAGCGAGGAGTTCGCCGAT---CTAAGGGGTAGATACGGTGTTGACGATTACTCTGAG 172723

Qy   1477 gtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaagcggctc 1536

Db 172724 CTTATGAAGGTTGTTGACGAGGCCCTCTCGGACTTCCCACAGCTTGACCCGCAGAGGGGT 172783

Qy   1537 ggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcagacgaac 1596

Db 172784 GGTGTAACGGGGGGGAGCTACGGGGGCTACATGACCAACGTTATAGTAACCAAAACTAGA 172843

Qy   1597 cgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcacggcgtc 1656

Db 172844 AGGTTCAAGGCGGCCGTCACACAGAGGAGCTGCAGCAACTGGATAAGCTTCTATGGCGAG 172903

Qy   1657 agtgatatcg---gctatttctttacagactggcagcttgagcatgacatgtttgaggac 1713

Db 172904 AGCGACATAGGATGGTATTTCGCCCCGGAACTTATCTCCGCGCAAGAACCATGGAGAGAC 172963

Qy   1714 acagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccgctt 1773

Db 172964 CTCGAGAAATACGTTGACTTTAGCCCGCTCTTCTCTGTTGAGAATATTGAGACACCACTG 173023

Qy   1774 ttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctgttt 1833

Db 173024 CTTATAATACACTCCACAGAGGACTTTAGATGCCCCCTCAGCGGTGCTATTCAACTCTTC 173083

Qy   1834 atcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcgcac 1893

Db 173084 ACCGCTCTAAAGCTGAAGGGTGTGGAGACCAGGCTTCTAGTGTTTCCAGGTGAGAACCAT 173143

Qy   1894 aatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagctca 1953

Db 173144 AACCTCTCCCGAAGCGGAACTCCTAAGAGGCGGGTGGCAAGGCTTAGGGCGATAGCCTCC 173203

Qy   1954 tggtttgatcaacatct 1970

Db 173204 TGGTTCAAGGAACACCT 173220
```

```
RESULT   9
CNSPAX05/c
LOCUS       CNSPAX05    304290 bp    DNA              BCT       07-AUG-2000
DEFINITION  Pyrococcus abyssi complete genome; segment 5/6.
ACCESSION   AJ248287 AL096836
VERSION     AJ248287.1  GI:5458657
KEYWORDS    complete genome.
SOURCE      Pyrococcus abyssi.
  ORGANISM  Pyrococcus abyssi
            Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
            Pyrococcus.
REFERENCE   1  (bases 1 to 304290)
  AUTHORS   Heilig,R.
  TITLE     Pyrococcus abyssi genome sequence: insights into archaeal
            chromosome structure and evolution
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 304290)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .304290
                     /organism="Pyrococcus abyssi"
                     /strain="Orsay"
                     /db_xref="taxon:29292"
     CDS             complement(77. .2398)
                     /note="PAB1560"
                     /codon_start=1
                     /transl_table=11
                     /product="hydrogenase expression/formation regulatory
                     protein (hypF)"
                     /protein_id="CAB50145.1"
                     /db_xref="GI:5458658"
                     /db_xref="SPTREMBL:Q9UZB4"
                     /translation="MKAYRIHVQGIVQAVGFRPFVYRIAHAHNLRGYVRNLGDAGVEI
                     VVEGREEDIEGFLRDLYRKKPPIAKIEKVEREEIPIQGFDRFYIEKSSTEKKGEGDSI
                     IPPDIAICEDCLRELFDPTNKRYMYPFIVCTNCGPRFTIIEDLPYDRENTAMREFPMC
                     DFCKSEYEDPLNRRYHAEPVACPVCGPSYRLYTSDGKEITGDPLRKAAKLIDKGYIIA
                     IKGIGGIHLACDATREDVVAELRKRLFRPQKPFAIMAKDLETVKGFAYISREEEEELT
                     SYRRPIVALRKREPFPLPENLAPGLHTIGVMLPYAGTHYILFHWSKTPVYVMTSANFP
                     GMPMIKDNEEAFEKLKDVADYLLLHNRRIPNRADDSVVRFVDGRRAVIRRSRGFVPLW
                     IEIPFEYKGLAVGAELMNAFGVAKNGKVYPSQYIGNTSKIEVLEFMREAIKHFFKILR
                     INALDLVVADLHPNYNTTKLAMEIAEEFNAEFLQVQHHYAHVASVMAEHNLEEVVGIA
                     LDGVGYGTDGKTWGGEIIYLSYEDIERLAHLEYYPLPGGDLASYYPLRALIGILSLNH
                     EIEEVENTIRKFCPNAIKSLKYGETELKVILRQLSSGINVAYASSTGRVLDAFSVLLN
                     VSYRRHYEGEPAMKLESFASQGKNDLKLTIPVEGEEIRVSELFEEVLDLMGKASPRDI
                     AYSIHLALARVFAEVSVEKAKEFGAKTVVLGGGVGYNELIVKTVRKIVEGNGLKFLTT
                     YEVPRGDNGINVGQAFLGGLYLEGYLNREDLSI"
     CDS             complement(2490. .3593)
                     /note="PAB1559"
                     /codon_start=1
                     /transl_table=11
                     /product="hydrogenase expression/formation protein (hypD)"
                     /protein_id="CAB50146.1"
                     /db_xref="GI:5458659"
                     /db_xref="SPTREMBL:Q9UZB3"
                     /translation="MLEKFRDKEIAQKIVRKIHKEAKGFDELRFMHVCGTHEDTVTRA
                     GIRSLLPENVKIMSGPGCPVCITPVEDIVRMMEIMRKAKEEGEDIILTTFGDMYRIPT
                     PMGSFADLKSEGFDVRVVYSIYDSYKIAQENRDKLVVHFSPGFETTTAPTAGMLETVI
                     NEGLDNFKIYSVHRLTPPAVEALLKAGTVFHGLIDPGHVSTIIGVRGWAYLTEKYNIP
                     QVVAGFEPVDVLLGILILIRMVKEGEAKIINEYSRVVKWEGNVKAQELINKFFEVKDA
                     KWRALGIIPQSGLELKREWKDLEIRTLYNPDVPKLPDLEKGCLCGAILRGLALPPQCP
                     HFGKTCTPRHPIGPCMVSYEGTCSIFYKYGALF"
     CDS             complement(3594. .3821)
                     /note="PAB7315"
                     /codon_start=1
                     /transl_table=11
                     /product="hydrogenase expression/formation protein (hypC)"
                     /protein_id="CAB50147.1"
                     /db_xref="GI:5458660"
                     /db_xref="SPTREMBL:Q9UZB2"
                     /translation="MCLAIPGRVVEIRDNVGIVDFGGVKREVRLDLLSDVKVGDYVIV
                     HTGFAIEKLDEERAKEILEAWDEVFSILGGE"
     CDS             3949. .4386
                     /note="PAB0821"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB50148.1"
                     /db_xref="GI:5458661"
                     /db_xref="SPTREMBL:Q9UZB1"
                     /translation="MQDSDLLSEIQELKRNLQVAFEVLAEIAQSYIRLLNLYAEYGGI
                     GIDVAIPEIRNDPIAREIVKILFDLRRANVSQITRELKGRRGKASRNTVRAKLNELKN
                     LGVVVEVPGERGKVYTLSKDVVKRWLDLIGIPIKLEHIKDYLR"
     CDS             4392. .4856
                     /note="PAB0822"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB50149.1"
                     /db_xref="GI:5458662"
                     /db_xref="SPTREMBL:Q9UZB0"
                     /translation="MEKEKLEKVVEELIEEIKSAKTEEEVEVLERKVKAVKKLIEAYE
                     GDKDLDKIAMLMDKLGPMVEDILEPLQKLLADLYSPEKMQALGKSVAQFYKDLVEAGM
                     DKETATELTKEFMKNINVMRSLLEGLTGMLQKGKGFQVYVHGKEEKREETKE"
     CDS             complement(5039. .5668)
                     /note="PAB1558"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB50150.1"
                     /db_xref="GI:5458663"
                     /db_xref="SPTREMBL:Q9UZA9"
                     /translation="MLKEIFNGIKFGEIALIEYDSNTTPVYVLHSLLSWSEEKGYKIV
                     IFDVFDSLIVYRKMAELLGFDTSLCKKAKVVKVGGRVNEGNVVMKLTVTEYGPFEKVL
```

EEAIEQVYEDKTIVIPLGTEKVLSLYPFREQLSFTNFLALRVGDKRKKAFHFINTDLV
KAVAPQILPMLEESFTTVMRLKKYGKVEKLIVVKSLNPKLDGLEVLAKS"
CDS complement(5709. .6227)
/note="PAB1557"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50151.1"
/db_xref="GI:5458664"
/db_xref="SPTREMBL:Q9UZA8"
/translation="MEGYRLLYPMRTYLVVSGHGDEANVMAADWVTVLSFDPFMVGVA
IAPKRTTHKLIKKYNEFVISVPSLEMLRDVWIAGTKKGPSKLKDMGITLVPSRKVKVP
SIKEALANIECEVYDTKDYGDHTLFVGKVLAYTYKEEAFEGGKPNLKFNFLAHVSWSE
FVTFQDKIYRAE"
CDS complement(6269. .6913)
/note="PAB1556"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50152.1"
/db_xref="GI:5458665"
/db_xref="SPTREMBL:Q9UZA7"
/translation="MSVEFSTMILLIMVLAAVSTLQFYKGRKLNLLLMQHYLRSIEAV
VKPKDKDYVWLGGYVGFRAYYKLEDKDILKFEYTLTLLPRQSILYFPISLLINRHDKL
YVVVRPTFDIKREVHFLQKGYFRMKPRIEREIELLKGEEEINGVKFEVLYEKSRDVEA
CDS VKEFISGFSKVKNVKHVALVPKTNVIYVFFKPEPESIEDDVRNIIRFVKRMANR"
complement(7184. .8176)
/note="PAB1555"
/codon_start=1
/transl_table=11
/product="ANION TRANSPORTING ATPASE"
/protein_id="CAB50153.1"
/db_xref="GI:5458666"
/db_xref="SPTREMBL:Q9UZA6"
/translation="MREFLTPKEGFRVIFVIGKGGVGKTTSSASISVALAKAGYRTLI
VSLDPAHNLGDVFMEKLSDKPREIIDNLYASELDMEGMIKGYLEHLEKNLKNMYRYLT
VINLEKYFEVLRYSPGIEEYATLEAIREILVKGDEWDVIVFDTPPTGLTLRVLALPRI
SLVWTEKLIDIRKKILERRRAITKIQGEQKFVIDGEEIKLPTREEEDAVMRELKAYRD
EIKFVEDVLTNPNKTSVIAVMNPEFLPLYETKRAYESLKKFKIPFNMVVMNKVIKLKR
CDS EIPEIRVKLEAQEKVLSEVKREFPGVEIVEIPMFQEEPRGIEWLSKVGEMIVGR"
complement(8489. .10267)
/note="PAB1554"
/codon_start=1
/transl_table=11
/product="CARBON STARVATION PROTEIN A."
/protein_id="CAB50154.1"
/db_xref="GI:5458667"
/db_xref="SPTREMBL:Q9UZA5"
/translation="MNSTVVILAAAIIYVGLYFTYGKGLQNKVVRADPNRPTPAHRLY
DGVDYVPAHPLVLYGHHFASIAGAGPIVGPAVAMAWGWLPSLIWVWFGNVFIGAVHDY
LALMSSVRYDGKSIQWIAGKLMSKRTSMAFELYVWFALILVIAAFAAVIAGIFVKTPG
AATASILFLLIAVLLGWLLYKVQINFKVGTIIGIILLIIAIYIGFKFPLVASAKAWYI
FLGAYVIVASSLPVWVLLQPRDYLNAYILWFGLVLGGLAFFIIGFKGVGTFTAPAYTT
FSAHVVGGKPSPFWPTIPLVIACGALSGFHSLVGSGTSSKQLDNELHGLLVGYGGMFT
EGFLSTIVISSIAVYGLQIFHNAGIGITAETWAQAYAPLVAAKSITTETLQKLGVAAL
ENFKNGTVIGKVGIFASSYAHGLHDAFGIDIHLGTLFASLWVSAFALTTLDTATRLGR
FAWQEIIEMVKGPEILKNKWVASFILVFFGIALAWKGKWLVIWPAFSGMNQMLASIAM
MTASLWVAKIQRPSGFWKWAVLIPALFLWVTVTSALIWFLIYVGKSSIWVSLITIVGL
VLNFLLVIDWYNAWKKPAEEYQAAAA"
CDS 11287. .13068
/note="PAB0824"
/codon_start=1
/transl_table=11
/product="RNASE L INHIBITOR"
/protein_id="CAB50155.1"
/db_xref="GI:5458668"
/db_xref="SPTREMBL:Q9UZA4"
/translation="MVRKMRIAVIDYDKCNPDKCGHFLCERVCPVNRMGGEAIIIDEE
NYKPIIQEASCTGCGICVHKCPFNAISIVNLPEQLDEDCVHRYGVNAFVLYRLPIVKD
GMVVGIVGPNGTGKTTAVKILAGQLIPNLCEDNDSWDNVIRAFRGNELQNYFERLKNG
EIRPVVKPQYVDLLPKAVKGKVRELLKKVDEVGKFEEVVKELELENVLDRELHQLSGG
ELQRVAIAAALLRKAHFYFFDEPSSYLDIRQRLKVARVIRRLANEGKAVLVVEHDLAV
LDYLSDVIHVVYGEPGVYGIFSKPKGTRNGINEFLQGYLKDENVRFRPYEIRFTKLSE
RVDVERETLVEYPRLVKDYGSFKLEVEPGEIRKGEVIGIVGPNGIGKTTFVKMLAGVE
EPTEGKVEWDLTVAYKPQYIKAEYEGTVYELLSKIDSSKLNSNFYKTELLKPLGIIDL
YDRNVEDLSGGELQRVAIAATLLRDADIYLLDEPSAYLDVEQRLAVSRAIRHLMEKNE
KTALVVEHDVLMIDYVSDRLIVFEGEPGRHGRALPPMGMREGMNRFLASVGITFRRDP
DSGRPRANKEGSVKDREQKARGEYYYA"
CDS 13130. .13513
/note="PAB0825"
/codon_start=1
/transl_table=11
/product="TRANSLATION INITIATION INHIBITOR, PUTATIVE."
/protein_id="CAB50156.1"
/db_xref="GI:5458669"
/db_xref="SPTREMBL:Q9UZA3"
/translation="MAKEVIFTEKAPKPIGPYSQAIKVGNFIFVAGQIPIDPETGEIV
KGDIKEQTKRVIENIKAILEEAGASLNDVVKVTVYLKDLNDFAKMNEVYSEYFGESKP
ARVAVEVSRLPKDVLIEMEAIAYKE"
CDS 13623. .14480
/note="PAB0826"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50157.1"

Query Match 8.0%; Score 158; DB 2; Length 304290;
Best Local Similarity 52.7%; Pred. No. 1.9e-37;
Matches 391; Conservative 0; Mismatches 345; Indels 6; Gaps 2;

```
Qy   1230 cggcgtgatggtgaacggctggctgatgaggcctgcacaaatggaaggtgagacaacata 1289
          || | ||  | |  ||||||||   || | ||| |  |  | ||| || ||||   |
Db 209349 CGACTTGGAGATAGACGGCTGGTACATAAAGCCCGAGCTCAAGGAGGGCGAGAAGGC--- 209291

Qy   1290 tccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttca 1349
          |||  | ||  |   | | ||||| || ||  |    ||||||||  |   |       |
Db 209292 TCCGGTAATAGTCTTCGTCCACGGTGGGCCCAAGGGAATGTACGGTTACTACTTCAAGTA 209233

Qy   1350 tgagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaag 1409
           ||| | ||| |  |||||  || || |||    | || ||| |||||||| || |||||
Db 209232 CGAGATGCAGCTTATGGCGAGCAAGGGCTACTACATAGTCTATGTCAATCCAAGGGGAAG 209173

Qy   1410 ccacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggatta 1469
          | |||| ||| |  |||| ||||    |  |||  |       |||   | |||| |
Db 209172 CAACGGTTACAGCGAGGACTTTGCACTTAGAGTCCTCGAGAGGACTGGCCTAGAGGACTT 209113

Qy   1470 tgacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaa 1529
           | ||| |  || | |   | || |||     |||| |  || |||||    ||     |
Db 209112 CCAGGATATACTGAATGGAATAGAGGAGTTCTTCAAGCTTGAACCGCAAGCGGACAGGGA 209053

Qy   1530 gcggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggca 1589
          | || | ||  | |||||   |||||||||||| || ||||| ||||||  | |   |||
Db 209052 GAGGGTTGGAATAACGGGAATAAGCTACGGAGGCTTCATGACGAACTGGGCCCTAACGCA 208993

Qy   1590 gacgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttca 1649
          ||   |||  || || || |   | | | ||  |   |||    | ||| | |    | |
Db 208992 GAGCGACCTATTCAAGGCCGGAATCAGCGAGAACGGAATAAGCTACTGGCTAACGAGTTA 208933

Qy   1650 cggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttga 1709
          ||   ||   || || || |  |  ||    |||  | |  ||     ||| |
Db 208932 CGCGTTCTCCGACATAGGATTGTGGTT---CGACAAGGAAGTTATAGGTGAGAACCCGCT 208876

Qy   1710 ggacacagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacacc 1769
          || |  || ||  || || | |    |||  |    |||||    |||||| || | ||
Db 208875 AGAGAACGAGAACTTCAGGAAGCTAAGTCCGCTGTTCTACGCTAAGAACGTGAAGGCTCC 208816

Qy   1770 gcttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagct 1829
          || ||| || | ||  ||      ||| | || || ||  | || |||     | | |
Db 208815 CCTGTTGCTAATCCACAGCCTTGAAGATTATCGCTGTCCATTGGATCAGAGCTTGATGTT 208756

Qy   1830 gtttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatc 1889
          |    ||  |||||| |  | ||||| |||    |  | |    ||   ||  | | |
Db 208755 CTACCACGTCCTGAAAGATTTAGGCAAAGAAGTTTACATAGCGATATTCAAGAGGGGAGC 208696

Qy   1890 gcacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcag 1949
          ||             |  |||  |||||||||  |||| ||| |    ||     |||
```

```
Db 208695 TCATGGACATAGCATAAGGGGAAGCCCAAGACACAGGATGAAGAGGTACAAGCTCTTCAT 208636

Qy   1950 ctcatggtttgatcaacatctc 1971
             |  || ||     | |||
Db 208635 GGAGTTCTTCGAGAGGAAGCTC 208614


RESULT 10
AX041921/c
LOCUS       AX041921   349980 bp    DNA              PAT       23-NOV-2000
DEFINITION  Sequence 816 from Patent WO0065062.
ACCESSION   AX041921
VERSION     AX041921.1  GI:11340686
KEYWORDS    .
SOURCE      Pyrococcus abyssi.
  ORGANISM  Pyrococcus abyssi
            Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
            Pyrococcus.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Forterre,P., Thierry,J.C., Prieur,D., Dietrich,J., Lecompte,O.,
            Querellou,J., Weissenbach,J., Saurin,W., Heilig,R., Flament,D.,
            Raffin,J.P., Henneke,G., Gueguen,Y. and Rolland,J.L.
  TITLE     Genome sequence and polypeptides of pyrococcus _i(abissy), fragment
            and uses thereof
  JOURNAL   Patent: WO 0065062-A 816 02-NOV-2000;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IFREMER
            INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Pyrococcus abyssi"
                     /db_xref="taxon:29292"
                     /note="Original length of seq 1: 1765118 (for 812
                     sequences) New sequences : seq 001, from 0.000.001 to
                     0.349.980, length: 349.980 seq 813, from 0.300.001 to
                     0.649.980, length: 349.980 seq 814, from 0.600.001 to
                     0.949.980, length: 349.980 seq 815, from 0.900.001 to
                     1.249.980, length: 349.980 seq 816, from 1.200.001 to
                     1.549.980, length: 349.980 seq 817, from 1.500.001 to
                     1.765.118, length: 265.118"
BASE COUNT    98084 a  80447 c  77665 g  93784 t
ORIGIN


  Query Match             8.0%;  Score 158;  DB 9;  Length 349980;
  Best Local Similarity   52.7%;  Pred. No. 1.9e-37;
  Matches  391;  Conservative    0;  Mismatches  345;  Indels    6;  Gaps    2;

Qy    1230 cggcgtgatggtgaacggctggctgatgaggcctgcacaaatggaaggtgagacaacata 1289
           || | ||  | |  |||||||   || | ||| |  |  | ||| || ||||   |
Db  205109 CGACTTGGAGATAGACGGCTGGTACATAAAGCCCGAGCTCAAGGAGGGCGAGAAGGC--- 205051

Qy    1290 tccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttca 1349
           |||  | ||  |   | | ||||| || ||  |    ||||||||  |    |      |
Db  205052 TCCGGTAATAGTCTTCGTCCACGGTGGGCCCAAGGGAATGTACGGTTACTACTTCAAGTA 204993

Qy    1350 tgagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaag 1409
            ||| | ||| |  |||||   || || |||    | || ||| |||||||| || |||||
Db  204992 CGAGATGCAGCTTATGGCGAGCAAGGGCTACTACATAGTCTATGTCAATCCAAGGGGAAG 204933

Qy    1410 ccacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggatta 1469
           | |||| ||| |  |||| ||||    |   |||  |        |||   | |||| |
Db  204932 CAACGGTTACAGCGAGGACTTTGCACTTAGAGTCCTCGAGAGGACTGGCCTAGAGGACTT 204873

Qy    1470 tgacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaa 1529
            | ||| |  || | |   | || |||    |||| |  || |||||    ||     |
Db  204872 CCAGGATATACTGAATGGAATAGAGGAGTTCTTCAAGCTTGAACCGCAAGCGGACAGGGA 204813

Qy    1530 gcggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggca 1589
           | || | ||  | |||||  ||||||||||| || ||||| ||||||  | |  ||| |
Db  204812 GAGGGTTGGAATAACGGGAATAAGCTACGGAGGCTTCATGACGAACTGGGCCCTAACGCA 204753

Qy    1590 gacgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttca 1649
           ||   |||  || || || |   | | | ||  |   |||    | ||| | |    | |

Db 204752 GAGCGACCTATTCAAGGCCGGAATCAGCGAGAACGGAATAAGCTACTGGCTAACGAGTTA 204693

Qy    1650 cggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttga 1709
           ||   ||   || || || |  |  ||    |||  | |  ||    ||| |
Db  204692 CGCGTTCTCCGACATAGGATTGTGGTT---CGACAAGGAAGTTATAGGTGAGAACCCGCT 204636

Qy    1710 ggacacagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacacc 1769
           || |  || ||  || || | |    |||  |    |||||    |||||| || | ||
Db  204635 AGAGAACGAGAACTTCAGGAAGCTAAGTCCGCTGTTCTACGCTAAGAACGTGAAGGCTCC 204576

Qy    1770 gcttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagct 1829
           || ||| || | ||  ||      ||| | || || ||  | || |||     |  | |
Db  204575 CCTGTTGCTAATCCACAGCCTTGAAGATTATCGCTGTCCATTGGATCAGAGCTTGATGTT 204516

Qy    1830 gtttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatc 1889
            |    ||  |||||| |  | ||||| |||    |  | |     ||  ||  | | |
Db  204515 CTACCACGTCCTGAAAGATTTAGGCAAAGAAGTTTACATAGCGATATTCAAGAGGGGAGC 204456

Qy    1890 gcacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcag 1949
            ||                |  |||  ||||||||| |||| ||| |    ||     |||
Db  204455 TCATGGACATAGCATAAGGGGAAGCCCAAGACACAGGATGAAGAGGTACAAGCTCTTCAT 204396

Qy    1950 ctcatggtttgatcaacatctc 1971
             |  || ||     | |||
Db  204395 GGAGTTCTTCGAGAGGAAGCTC 204374


RESULT 11
AP000002
LOCUS       AP000002   257000 bp    DNA              BCT       06-APR-2000
DEFINITION  Pyrococcus horikoshii OT3 genomic DNA, 287001-544000 nt. position
            (2/7).
ACCESSION   AP000002 AB009475 AB009476 AB009477 AB009478 AB009479 AB009480
            AB009481 AB009482 AB009483 AB009484 BA000001
VERSION     AP000002.1  GI:3236129
KEYWORDS    .
SOURCE      Pyrococcus horikoshii (strain:OT3) DNA.
  ORGANISM  Pyrococcus horikoshii
            Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
            Pyrococcus.
REFERENCE   1  (sites)
  AUTHORS   Kawarabayasi,Y., Sawada,M., Horikawa,H., Haikawa,Y., Hino,Y.,
            Yamamoto,S., Sekine,M., Baba,S., Kosugi,H., Hosoyama,A., Nagai,Y.,
            Sakai,M., Ogura,K., Otuka,R., Nakazawa,H., Takamiya,M., Ohfuku,Y.,
            Funahashi,T., Tanaka,T., Kudoh,Y., Yamazaki,J., Kushida,N.,
            Oguchi,A., Aoki,K., Nakamura,Y., Robb,T.F., Horikoshi,K.,
            Masuchi,Y., Shizuya,H. and Kikuchi,H.
  TITLE     Complete sequence and gene organization of the genome of a
            hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3
  JOURNAL   DNA Res. 5 (2), 55-76 (1998)
  MEDLINE   98344137
REFERENCE   2  (bases 1 to 257000)
  AUTHORS   Tanaka,T., Kawarabayasi,Y. and Kikuchi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) to the DDBJ/EMBL/GenBank databases. Yutaka
            Kawarabayasi, National Institute of Technology and Evaluation,
            Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo
            151-0066, Japan (E-mail:genomeOT3@nite.go.jp, Tel:+81-3-3481-8951,
            Fax:+81-3-3481-8424)
  COMMENT   On Jul 28, 1998 this sequence version replaced gi:3130553
            gi:3130577 gi:3130627 gi:3130656 gi:3130685 gi:3130705 gi:3130737
            gi:3130785 gi:3130824 gi:3130828.
            AB009475-AB009484: submitted (10-DEC-1997)
            Kawarabayasi, Y. is officially affiliated with the National
            Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
            305-0046, Japan.
            Robb, T. F. is at the Center of Marine Biotechnology, University of
            Maryland, Baltimore, MD, USA.
            Horikoshi, K. is at the Japan Marine Science and Technology Center,
            Yokosuka, Kanagawa 237-0061, Japan.
            Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
            Japan.
            Shizuya, H. is at the California Institute of Technology, Pasadena,
```

CA, USA.
The other authors are at the National Institute of Technology and Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against GenBank and GenPept release 103; EMBL release 52.0; SwissProt release 34.0; PIR-Protein release 54.0; and OWL release 29.5.
E-mail address for comments and questions: genomeOT3@nite.go.jp
Restriction map, ORF organization, sequence alignment and more information are available at W.W.W. site of Biotechnology Center, URL: http://www.bio.nite.go.jp/.

FEATURES Location/Qualifiers

source 1. .257000
/organism="Pyrococcus horikoshii"
/strain="OT3"
/db_xref="taxon:53953"
/map="287001-544000 nt. position"

gene 1363. .2664
/gene="PH0318"

CDS 1363. .2664
/gene="PH0318"
/note="similar to PIR:E64485 percent identity: 55.181 in 403aa; PIR:S74842 percent identity: 49.869 in 397aa; Swiss_Prot:P33221 percent identity: 50.773 in 404aa"
/codon_start=1
/transl_table=11
/product="433aa long hypothetical phosphoribosylglycinamide formyl transferase"
/protein_id="BAA29392.1"
/db_xref="GI:3256709"
/translation="MVVMIKLRDELGTATTDSAQKILLLGSGELGKEIAIEAQRLGVEVVAVDRYANAPAMQVAHRSYVGNMMDKDFLWSVVEREKPDAIIPEIEAINLDALFEFEKDGYFVVPNARATWIAMHRERLRETLVKEAKVPTSRYMYATTLDELYEACEKIGYPCHTKAIMSSSGKGSYFVKGPEDIPKAWEEAKTKARGSAEKIIVEEHIDFDVEVTELAVRHFDENGEIVTTFPKPVGHYQIDGDYHASWQPAEISEKAEREVYRIAKRITDVLGGLGIFGVEMFVKGDKVWANEVSPRPHDTGMVTLASHPPGFSEFALHLRAVLGLPIPGEWVDGYRLFPMLIPAATHVIKAKVSGYSPRFRGLVKALSVPNATVRLFGKPEAYVGRRLGIALAWDKDVEVAKRKAEMVAHMIELRTRSSDWHDQNYEKRKHLLR"

gene complement(1645. .1974)
/gene="PH0319"

CDS complement(1645. .1974)
/gene="PH0319"
/codon_start=1
/transl_table=11
/product="109aa long hypothetical protein"
/protein_id="BAA29393.1"
/db_xref="GI:3256710"
/translation="MFLNYYLLCTTSRFSLCFFPGFRNVLGPLDEVRSLSRGAHYSLCVAWITDLFTGLIEFIKGCSVHIPGSRNLCFLHKGFPKPLPVHSYPGSSCIWNYKITILEFEEGI"

gene complement(2653. .3186)
/gene="PH0320"

CDS complement(2653. .3186)
/gene="PH0320"
/note="similar to owl:SSOUREK percent identity: 68.000 in 150aa; PIR:S76001 percent identity: 60.694 in 173aa; owl:BAPURKE1 percent identity: 62.329 in 146aa"
/codon_start=1
/transl_table=11
/product="177aa long hypothetical phosphoribosylaminoimidazole carboxylase catalytic subunit"
/protein_id="BAA29394.1"
/db_xref="GI:3256711"
/translation="MVISMKSEKPLVGIIMGSDSDLPVMKEAARILEEFGVPYEITIISAHRTPERAYEYAKKAEERGIEVIIAGAGGAAHLPGIIASLTVLPVIGVPIKSKALNGLDSLLSIVQMPSGIPVATVAIDNAKNAALLALRILGIKYPEIKEKLRRYMKDMKRKVEEKAKRLEEMGWERYLSE"

gene complement(3275. .4036)
/gene="PH0321"

CDS complement(3275. .4036)
/gene="PH0321"
/codon_start=1
/transl_table=11
/product="253aa long hypothetical protein"
/protein_id="BAA29395.1"
/db_xref="GI:3256712"
/translation="MISWEIEELDREIGKIKKHSLILIHEEDASSRGKDILFYILSRKLKSDNLVGMFSISYPLQLIIRILSRFGVDVIKYLENHRLAIVDTFGSFHGIKATMPGVWYLEGMLSSETLPIKYAKAVEDHKKVWMDLNLFEGRELYGFAISMSGYLEVFTPEETLRYLETSAEVRYGHPAYKKYPRGTNFWLWEGVKDKRVLLSVYRRADYVLKTRSSLGENGIKRELLVIKTPKPIEELVRFEYEFKGNEPKLRRVD"

gene complement(4039. .4698)
/gene="PH0322"

CDS complement(4039. .4698)
/gene="PH0322"
/codon_start=1
/transl_table=11
/product="219aa long hypothetical protein"
/protein_id="BAA29396.1"
/db_xref="GI:3256713"
/translation="MGIIEELVSSIPEGGILAIIQKELESEGDRFGLLVLKHLLENENNVFLFLYEPLRTFMTNMENLGINVDEYLGENLMIFDVFGSMNKIERKIKGVYKLSGYLDDVVFVSKLKEWAIKTLKTTNIANFWFFTYLSSGICKLFSNPLLTYKLIWALREEILKGYRPKTIITYSQLECPILEETVYLASDIVLETRIINGKKVGIITKGPNENLIFELFKEV"

gene 4856. .6172
/gene="PH0323"

CDS 4856. .6172
/gene="PH0323"
/note="similar to PIR:A64417 percent identity: 55.478 in 442aa; owl:HUMRPS4 percent identity: 33.738 in 424aa; Swiss_Prot:P12039 percent identity: 33.819 in 352aa. motif=prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="438aa long hypothetical phosphoribosylamide--glycine ligase"
/protein_id="BAA29397.1"
/db_xref="GI:3256714"
/translation="MKVLLVGGGGREHAIGEALVKGGAELYVVSNHRNPGLMRLAKDYGLAKETNVEDVIKFARKWGIELAFIGPEAPLEAGIVNALEEEGIPAVGPTREAARLETNKAWAREFMERNNIPGRKMFRIFDDVQEMRKWIDEYGKPVVVKPLGLTGGKGVKVVGYQLKDNEEAKEYAEHIIRKDGKVLIEERTDGVEFTLQVFTDGKKVIPMPLVQDYPHAYEGDVGPITGGMGSYSCSNHLLPFITEGDFERALKTLEETIEAMRKEGYPYKGILYGQFMLSGEGPVLIEYNARFGDPEAINVLAVLDDNLLEIAKGIVEGSLRKAKFLNKATVVKYIAPQGYPQDPIKGIRIEVDEEGIKNEGAKIIYAAVDENLTLLGSRALAIVGVADSLEEAERIAENGVSYVKGPIFYRKDVGTRESVEKRIEIMKKLGKEFEPNLC"

gene 6174. .7325
/gene="PH0324"

CDS 6174. .7325
/gene="PH0324"
/note="similar to PIR:H64316 percent identity: 44.847 in 382aa"
/codon_start=1
/transl_table=11
/product="383aa long hypothetical protein"
/protein_id="BAA29398.1"
/db_xref="GI:3256715"
/translation="MIYMIDREEILSILEKYDKDKITIGVIGSHSALDIADGAKEEGFPVLVVAQRGRHRTYAKYFKLRKTKDNLTKGFIDEVIILDKFAQIIEIQEELIKRNVIFIPNRSFVVYTGIDKVENEFKVPMFGTRSLLRTEERREEKSYYWLLEKAGLPYPEEVKPEEIDDVGLVMVKLPHAKKRLERGFFTAASYKEFKEKSERLIRLGVITREDLDKARIERYIIGPVFNFDFFYSPIDEEIELLGVDWRFETSLDGHVRLPAAQQLTLPEHQFEPEYTVCGHASATLRESLLEKVFDMAEKYVEATKKYYPPGIIGPFTLQTVVDKDLNFYIFDVAPRTGGGTNIHMSMGHPYGNSLWRSPMSTGRRIALEIKRGIELDELEKLVT"

gene 7392. .8486
/gene="PH0325"

CDS 7392. .8486
/gene="PH0325"
/codon_start=1
/transl_table=11
/product="364aa long hypothetical protein"
/protein_id="BAA29399.1"

```
                     /db_xref="GI:3256716"
                     /translation="MSSAIKIAEEFYNDEYSDSVLYAELAKYEKNKEIKEEFLRLSRM
                     EAKHAKFWHDFLVRRGVKPKKPRIKKLTFFTVKLLRKLFGPAMVVSMLEMGENSAIQK
                     YFKFFGEFSDEITEEEFEGLRGIIIDELEHEKFFSESKKLFHVENIRDFVLGMNDGLV
                     EILGAVTGLSAVYPNSPRLVGISGLIVGVAGALSMGIGALISVRSQRQVSEAIRERTK
                     ILFKVSPEKAKEELYEKLVEGGLPEEIAREVSEKLLEKEEAIKLLVPEEEENEFRAA
                     LYTGIAYLFGVAFPVTPYFFASTSLRALPISVTLAGLALAIVATSISLISGISIRKKV
                     IEMVTTGLGAAFLSYIFGHIMESLFNVSAL"
     gene            8483..11494
                     /gene="PH0326"

Query Match              7.6%;  Score 150.6;  DB 2;  Length 257000;
Best Local Similarity   52.3%;  Pred. No. 3.9e-35;
Matches  383;  Conservative    0;  Mismatches  344;  Indels    6;  Gaps    2;

Qy     1239 ggtgaacggctggctgatgaggcctgcacaaatggaaggtgagacaacatatccacttat 1298
Db   247119 GCTTGACGGATGGTACATAAAGCCAGAGATCAAGGAGGGAGAGAAAGCA---CCAGTAAT 247175

Qy     1299 tcttaacatacacggcggtccgcatatgatgtacggacatacatattttcatgagtttca 1358
Db   247176 AGTCTTCGTCCACGGTGGGCCAAAGGGAATGTACGGCTATTACTTTAAGTACGAGATGCA 247235

Qy     1359 ggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagccacggcta 1418
Db   247236 ACTGATGGCGAGCAAAGGGTACTACATAGTTTACGTGAACCCAAGGGGAAGCAACGGTTA 247295

Qy     1419 cgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgatgt 1478
Db   247296 CAGCGAGGATTTTGCCCTGAGGGTTCTAGAGAGAACGGGACTTGAAGACTTCCAGGATAT 247355

Qy     1479 gatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcgg 1538
Db   247356 ATTGAATGGAATAGAGGAATTCCTCAGGCTAGAACCACAAGCTGACAGGGAGAGGATCGG 247415

Qy     1539 tgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcagacgaaccg 1598
Db   247416 TATCACGGGAATAAGCTACGGTGGCTACATGACGAACTGGGCCTTAACCCAGAGTGACCT 247475

Qy     1599 ctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcacggcgtcag 1658
Db   247476 CTTTAAAGCTGGAATAAGCGAGAACGGAATAAGCTACTGGTTAACAAGTTACGCCTTCTC 247535

Qy     1659 tgatatcggctatttctttacagactggcagcttgagcatgacatgtttgaggacacaga 1718
Db   247536 CGATATAGGGTTATGGTT---CGACAAGGAGGTTATTGGTGACAACCCGCTTGAAAACGA 247592

Qy     1719 aaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccgcttttgat 1778
Db   247593 GAACTATAGAAAGTTAAGCCCCCTATTCTATGCTAAGAACGTTAAGGCCCCCTTACTACT 247652

Qy     1779 actgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctgtttatcgc 1838
Db   247653 GATCCACAGCCTTGAGGATTACAGGTGCCCATTGGATCAGAGCTTAATGTTCTATCACGT 247712

Qy     1839 tctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcgcacaattt 1898
Db   247713 CCTCAAGGATCTCGGTAAGGAAGTTTACATTGCGATATTCAAGAAGGGAGCGCATGGGCA 247772

Qy     1899 atcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagctcatggtt 1958
Db   247773 TAGCATAAGGGGATCTCCAAGGCACAGGATGAAGAGGTACAAGCTATTCATGGAGTTCTT 247832

Qy     1959 tgatcaacatctc 1971
Db   247833 TGAGAGGAAGCTC 247845

RESULT 12
AR086952
LOCUS       AR086952      1869 bp    DNA              PAT      07-SEP-2000
DEFINITION  Sequence 1 from patent US 5985646.
ACCESSION   AR086952
VERSION     AR086952.1  GI:10013718
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1869)
  AUTHORS   Murphy,D., Reid,J. and Robertson,D.
  TITLE     Amidase
  JOURNAL   Patent: US 5985646-A 1 16-NOV-1999;
FEATURES             Location/Qualifiers
     source          1..1869
                     /organism="unknown"
BASE COUNT      476 a    486 c    567 g    340 t
ORIGIN

Query Match              7.6%;  Score 149.8;  DB 9;  Length 1869;
Best Local Similarity   52.3%;  Pred. No. 2e-35;
Matches  356;  Conservative    0;  Mismatches  322;  Indels    3;  Gaps    1;

Qy       1291 ccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttcat 1350
Db       1135 CCGGTGATAGTCTTCGTCCACGGCGGGCCGAAGGGCATGTACGGACACCGCTTCGTCTAC 1194

Qy       1351 gagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagc 1410
Db       1195 GAGATGCAGCTGATGGCGAGCAAGGGCTACTACTGCTGCTTCGTGAACCCGCGCGGCAGC 1254

Qy       1411 cacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattat 1470
Db       1255 GACGGCTATAGCGAAGACTTCGCGCTCCGCGTCCTGGAGAGGACTGGCTTGGAGGACTTT 1314

Qy       1471 gacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaag 1530
Db       1315 GAGGACATAATGAACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCAGGCCGACAGGGAG 1374

Qy       1531 cggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcag 1590
Db       1375 CGCGTTGGAATAACGGGCATAAGCTACGGCGGCTTCATGACCAACTGGGCCTTGACTCAG 1434

Qy       1591 acgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcac 1650
Db       1435 AGCGACCTCTTCAAGGCAGGAATAAGCGAGAACGGCATAAGCTACTGGCTCACCAGCTAC 1494

Qy       1651 ggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttgag 1710
Db       1495 GCCTTCTCGGACATAGGGCTCTGGTACGACGTCGAGGTCATCGGGCCAAATCCGTTAGAG 1554

Qy       1711 gacacagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccg 1770
Db       1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTGTTCTACGCTCAGAACGTGAAGGCGCCG 1611

Qy       1771 cttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctg 1830
Db       1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCCGCTCGACCAGAGCCTTATGTTC 1671

Qy       1831 tttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcg 1890
Db       1672 TACAACGTGCTCAAGGACATGGGCAAGGAAGCCTACATAGCGATATTCAAGCGCGGCGCC 1731

Qy       1891 cacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagc 1950
Db       1732 CACGGCCACAGCGTCCGCGGAAGCCCGAGGCACAGGCCGAAGCGCTACAGGCTCTTCATA 1791

Qy       1951 tcatggtttgatcaacatctc 1971
Db       1792 GAGTTCTTCGAGCGCAAGCTC 1812

RESULT 13
AR095647
LOCUS       AR095647      1869 bp    DNA              PAT      08-SEP-2000
DEFINITION  Sequence 1 from patent US 6004796.
```

```
ACCESSION   AR095647
VERSION     AR095647.1  GI:10023708
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1869)
  AUTHORS   Murphy,D., Reid,J. and Robertson,D.
  TITLE     Amidase
  JOURNAL   Patent: US 6004796-A 1 21-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .1869
                     /organism="unknown"
BASE COUNT      476 a    486 c    567 g    340 t
ORIGIN
```

Query Match 7.6%; Score 149.8; DB 9; Length 1869;
Best Local Similarity 52.3%; Pred. No. 2e-35;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

```
Qy   1291 ccacttattcttaacatacacggcggtccgcatatgatgtacggacatacatattttcat 1350
          ||  | ||  |   | | |||||||| ||| |    ||||||||||||    |   | |
Db   1135 CCGGTGATAGTCTTCGTCCACGGCGGGCCGAAGGGCATGTACGGACACCGCTTCGTCTAC 1194

Qy   1351 gagtttcaggtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagc 1410
          ||| | ||| || |||||   || || |||     |   |   | || ||| | || |||
Db   1195 GAGATGCAGCTGATGGCGAGCAAGGGCTACTACTGCTGCTTCGTGAACCCGCGCGGCAGC 1254

Qy   1411 cacggctacgggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattat 1470
          |||||||  |  | || || | |      |||  |      |||    |||| | |
Db   1255 GACGGCTATAGCGAAGACTTCGCGCTCCGCGTCCTGGAGAGGACTGGCTTGGAGGACTTT 1314

Qy   1471 gacgatgtgatgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaag 1530
          || ||  | ||| | |   | || |||    |||| |  || |||||    ||    ||
Db   1315 GAGGACATAATGAACGGCATCGAGGAGTTCTTCAAGCTCGAACCGCAGGCCGACAGGGAG 1374

Qy   1531 cggctcggtgtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcag 1590
          ||  | ||  | ||||||  |||||||||| || || ||||||||||||  | |    |||
Db   1375 CGCGTTGGAATAACGGGCATAAGCTACGGCGGCTTCATGACCAACTGGGCCTTGACTCAG 1434

Qy   1591 acgaaccgctttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcac 1650
          |   ||| ||| || || |   | | | ||  |   |||    | ||| ||| |    ||
Db   1435 AGCGACCTCTTCAAGGCAGGAATAAGCGAGAACGGCATAAGCTACTGGCTCACCAGCTAC 1494

Qy   1651 ggcgtcagtgatatcggctatttctttacagactggcagcttgagcatgacatgtttgag 1710
          | | ||   || || ||    | |     | | |    | | ||   |   ||| |||
Db   1495 GCCTTCTCGGACATAGGGCTCTGGTACGACGTCGAGGTCATCGGGCCAAATCCGTTAGAG 1554

Qy   1711 gacacagaaaagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccg 1770
          ||   || ||  || || | |    ||  |   |||||    |||||| || | |||
Db   1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTGTTCTACGCTCAGAACGTGAAGGCGCCG 1611

Qy   1771 cttttgatactgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctg 1830
          |  |  || | ||          |||  |||| || ||| |||| |||        ||
Db   1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCCGCTCGACCAGAGCCTTATGTTC 1671

Qy   1831 tttatcgctctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcg 1890
          | | ||  || ||  | |||||||||||| || |  | |     ||   |   |
Db   1672 TACAACGTGCTCAAGGACATGGGCAAGGAAGCCTACATAGCGATATTCAAGCGCGGCGCC 1731

Qy   1891 cacaatttatcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagc 1950
          |||           |  |||| ||| || ||  ||   ||||||   |      |||
Db   1732 CACGGCCACAGCGTCCGCGGAAGCCCGAGGCACAGGCCGAAGCGCTACAGGCTCTTCATA 1791

Qy   1951 tcatggtttgatcaacatctc 1971
             | || || |   | |||
Db   1792 GAGTTCTTCGAGCGCAAGCTC 1812
```

```
RESULT  14
AP000993
LOCUS       AP000993   343800 bp    DNA             BCT       23-FEB-2001
```

```
DEFINITION  Thermoplasma volcanium genomic DNA, complete sequence, section 3/6.
ACCESSION   AP000993 BA000011
VERSION     AP000993.1  GI:11990452
KEYWORDS    HTG.
SOURCE      Thermoplasma volcanium (strain:GSS1) DNA.
  ORGANISM  Thermoplasma volcanium
            Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
            Thermoplasma.
REFERENCE   1  (sites)
  AUTHORS   Kawashima,T., Yamamoto,Y., Aramaki,H., Nunoshiba,T., Kawamoto,T.,
            Watanabe,K., Yamazaki,M., Kanehori,K., Amano,N., Ohya,Y., Makino,K.
            and Suzuki,M.
  TITLE     Determination of the complete genomic DNA sequence of Thermoplasma
            volvanium GSS1
  JOURNAL   Proc. Jpn. Acad. 75, 213-218 (1999)
REFERENCE   2  (sites)
  AUTHORS   Kawashima,T., Amano,N., Koike,H., Makino,S., Higuchi,S.,
            Kawashima-Ohya,Y., Watanabe,K., Yamazaki,M., Kanehori,K.,
            Kawamoto,T., Nunoshiba,T., Yamamoto,Y., Aramaki,H., Makino,K. and
            Suzuki,M.
  TITLE     Archaeal adaptation to higher temperatures revealed by genomic
            sequence of thermoplasma volcanium
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (26), 14257-14262 (2000)
  MEDLINE   20570466
REFERENCE   3  (bases 1 to 343800)
  AUTHORS   Suzuki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-1999) Masashi Suzuki, National Institute of
            Bioscience and Human-Technology, Centre of Structural Biology;
            Higashi 1-1, Tsukuba, Ibaraki 305-0046, Japan
            (E-mail:namano@nibh.go.jp,
            URL:http://www.aist.go.jp/RIODB/archaic/, Tel:81-298-54-6582,
            Fax:81-298-54-6041)
COMMENT     The most recent annotation is available at
            http://www.aist.go.jp/RIODB/archaic/.
FEATURES             Location/Qualifiers
     source          1. .343800
                     /organism="Thermoplasma volcanium"
                     /strain="GSS1"
                     /db_xref="taxon:50339"
BASE COUNT    105792 a   65548 c   68174 g  104286 t
ORIGIN
```

Query Match 6.7%; Score 132.6; DB 2; Length 343800;
Best Local Similarity 51.5%; Pred. No. 2e-29;
Matches 365; Conservative 0; Mismatches 329; Indels 15; Gaps 2;

```
Qy   1269 aatggaaggtgagacaacatatccacttattcttaacatacacggcggtccgcatatgat 1328
          |||   |  || | ||  ||||   || ||     | ||||| || || ||    |
Db 306728 AATAATAACAGATAAAAATAATCCGACGATACTCTTTGTGCACGGTGGGCCTCACGCCAG 306787

Qy   1329 gtacggacatacatattttcatgagtttcaggtgctggcggcgaaaggatacgcggtcgt 1388
          || ||  || |||| |     || ||| |   |||||    || || | |   ||  |
Db 306788 TTATGGGAATGCATACTACATAGAATTTAACTACCTGGCCAGCAATGGCTTCAACGTTAT 306847

Qy   1389 ttatatcaatccgagaggaagccacggctacgggcaggaatttgtgaatgcggtcagagg 1448
          |     |||||    |||||   ||| || ||  | || || |       | || |  ||
Db 306848 ATTCGGAAATCCACATGGAAGTGGCGGTTATGGTGAAGATTTCGCTGCCTCTGTAATTGG 306907

Qy   1449 agattatgggggaaaggattatgacgatgtgatgcaggctgtggatgaggctatcaaacg 1508
          |||||  || || |||||||| || ||  | ||      |||    || ||||   |
Db 306908 AGATTGGGGTGGCAAGGATTACGAGGAACTTATATCATTTGTCAGAGATGCTAAGGATAA 306967

Qy   1509 agatccgcatattgatcctaagcggctcggtgtcacgggcggaagctacggaggttttat 1568
          | ||   |   | ||   |       ||   || || || ||  ||  || |||| |||
Db 306968 ATATGGCCTATTCGACAAT------TTCCACGTTACCGGAGGTTCTTATGGCGGTTATAT 307021

Qy   1569 gaccaactggatcgtcgggcagacgaaccgctttaaagctgccgttacccagcgctcgat 1628
          ||| ||    || ||     | ||| |    |||| | |||||||| |  || |    ||  |
Db 307022 GACAAATCATATTGTAACTAAAACGGATTTATTTAGATCTGCCGTTTCGGAGAGAAGCAT 307081

Qy   1629 atcaaattggatcagctttcacggcgtcagtgatatcggctatttctttacagactggca 1688
```

```
Db 307082 ATCGAATTTTTTGAGCATGTGCGGTACTAGCGATATAGGGTTTTGGTTCAATGCCATAGA 307141

Qy   1689 gctt---------gagcatgacatgtttgaggacacagaaaagctctgggaccggtctcc 1739

Db 307142 AAATGCCGTTAAGGATCCTTGGATATCTGAAAATCAAATTACACTGATGAAAATGTCACC 307201

Qy   1740 tttaaaatacgcagcaaacgtggagacaccgcttttgatactgcatggcgagcgggatga 1799

Db 307202 AATTTATCACGTAAAGAGAGTTAAAACGCCAATATTGTTTATACACGGAGAGAACGATTC 307261

Qy   1800 ccgatgcccgatcgagcaggcggagcagctgtttatcgctctgaaaaaaatgggcaagga 1859

Db 307262 CAGATGTCCTATAGAACAGGCCGAACAATTTTACACGGCTTTGAAGATAAACGGAGTGGA 307321

Qy   1860 aaccaagcttgtccgttttccgaatgcatcgcacaatttatcacgcaccggacacccaag 1919

Db 307322 AACAAAACTTGTCAGAAGTATCGGTGATTCACATGAACATGCCCGAAAAGGAAACCCAGA 307381

Qy   1920 acagcggatcaagcgcctgaattatatcagctcatggtttgatcaacat 1968

Db 307382 GAACATGAGAAAGAGGCTAGAACTCAAACTTGAATGGTTCAAATCACAT 307430


RESULT  15
AE005114
LOCUS       AE005114     9825 bp    DNA             BCT       12-FEB-2001
DEFINITION  Halobacterium sp. NRC-1 section 145 of 170 of the complete genome.
ACCESSION   AE005114 AE004437
VERSION     AE005114.1  GI:10581712
KEYWORDS    .
SOURCE      Halobacterium sp. NRC-1.
  ORGANISM  Halobacterium sp. NRC-1
            Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
            Halobacterium.
REFERENCE   1  (bases 1 to 9825)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     From the cover: genome sequence of halobacterium species NRC-1
  JOURNAL   Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)
   PUBMED   11016950
REFERENCE   2  (bases 1 to 9825)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Institute for Systems Biology, 4225
            Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES             Location/Qualifiers
     source          1. .9825
                     /organism="Halobacterium sp. NRC-1"
                     /strain="NRC-1"
                     /db_xref="taxon:64091"
     gene            262. .471
                     /gene="VNG2299H"
     CDS             262. .471
                     /gene="VNG2299H"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng2299h"
                     /protein_id="AAG20412.1"
                     /db_xref="GI:10581713"
                     /translation="MRTRRQFLATTTSLTTVGLLAGCARSPDNTDERGAENTYTASIA
                     PVGDVAFDGELFDRQRVADIVAGAI"
     gene            485. .1369
                     /gene="txrB3"
                     /note="VNG2301G"
     CDS             485. .1369
                     /gene="txrB3"
                     /note="TxrB3"
                     /codon_start=1
                     /transl_table=11
                     /product="thioredoxin reductase"
                     /protein_id="AAG20413.1"
                     /db_xref="GI:10581714"
                     /translation="MPTQDGERRDVVIVGGGPAGCAAGVFTARYGLDTVVFDRGNAAL
                     PRCAFVENYPGFPGGIDVPTLRGLFHDHAETAGCDLIADTVESVDRPSDDDTGFVVET
                     QDGRRVYTDTVLAAAWYDGSYLRPVVGDSAFETHDHHGESRERFDDAYADADGRTPVD
                     GLYVASPGGQRSAQAVIAAGNGAHVARCLLADRKRARGYPEGVAPHYDWKRRESDLSG
                     EWADRDRWREWFAAEAGDDHDLDDDEFAALRAAHLDRTFDATLSADAIEERAEAGAHR
                     LLDHIDDDHIESYREQRD"
     gene            1433. .3457
                     /gene="yuxL"
                     /note="VNG2302G"
     CDS             1433. .3457
                     /gene="yuxL"
                     /note="YuxL"
                     /codon_start=1
                     /transl_table=11
                     /product="acylaminoacyl-peptidase"
                     /protein_id="AAG20414.1"
                     /db_xref="GI:10581715"
                     /translation="MQPVEASDYWELARLSQPAVSPTGDRVAYVRTTAERDGDTGQET
                     ASVHVAPVGGGGSRRFTAADGTDAEPAWSPTGDRLAFVSTRGDAGGGRPQLWVVPVDG
                     GEATRITDVIGGVSNIAWGPDGDRVAFLQAVRPAERAEELDREHDGEYERDTPDPRVI
                     DRQMYRGHTSYRDGARRHVYVADLAAGRVDRVTAGVAECGGPAWGDDGTLYYPIRRGL
                     DADDRLEWAIEAATPADSADSTVVTTVEGMGPTLAVHGSRVAYTTTPADDPTLTQTAV
                     EVFDRDTGATDRLTDGVDRTVAERTAGHAPEWGPDGEHLYVCTPDEGGYVLRRVAADT
                     GDDRVVLAGDRHVHAFSVSADAVAVVQSEWDHPGEVVAATPMGGEEVTLTRANRDYLD
                     ERAVSKPESVWIDGPAGDIQGWVLTPPEFDPEETYPLVVNVHGGPHLMWSAAGTMWHE
                     FQTLAARGYVVFWCNPRGSTGYGQEWLRAVERDWGAVTLADVMAGVAAVTDRDYVDAD
                     NAFITGGSFGGFMTGWAVGHTDAFAGAVAQRGVYDLTGFYGTTDAYALVEGEFDTDPV
                     RDNAFLWAQSPAAHTDAVDTPTLLVHSEADYRTPANTAELFYRLLKKHGVDTRLVRYP
                     GEGHDLSRSGDPGHVVDRIERIAAWFDGYSDHHEAAPVLETDRPGAAADE"
     gene            complement(3469. .3708)
                     /gene="VNG2304H"
     CDS             complement(3469. .3708)
                     /gene="VNG2304H"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng2304h"
                     /protein_id="AAG20415.1"
                     /db_xref="GI:10581716"
                     /translation="MSDADAEPCDWCGDAVANPLSRTVRVSVDRSQIDSQRLCPECFA
                     EWVDRYETEMRADTGHDDAAAAADGEDGGTDIIVD"
     gene            complement(3755. .4438)
                     /gene="VNG2305C"
     CDS             complement(3755. .4438)
                     /gene="VNG2305C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng2305c"
                     /protein_id="AAG20416.1"
                     /db_xref="GI:10581717"
                     /translation="MTIEDRGDAHVITHSLAKHTLSRLRDTTTEQVSFRKGLVKLGRI
                     SGYEIIDGAMETEFTAFETPLTETTGERVTGLDDVVIINVLRAATPFVEGLLKAFPRA
                     KQGVISAGRDEDAGMDADGEFPITVDYVKLPDISEGDTVIVADPMLATGSTMCTVLDH
                     VTDEMPAAGVEDLFVLSAVSAPDGLLRVGDQFPDADLLTVAIDDTLTDDGYIVPGLGD
                     AGDRAFRTT"
     gene            4532. .5215
```

```
                     /gene="VNG2306H"
CDS                  4532. .5215
                     /gene="VNG2306H"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng2306h"
                     /protein_id="AAG20417.1"
                     /db_xref="GI:10581718"
                     /translation="MEESVAGFKQRGTWGDVVEHGERITTALREAGADEAFADAFAEW
                     DDWRPKSHERIGDEVNEKTADQASVAEGKGEKAGASADEDLQKAGEQLSASYEKLGDG
                     DGGDAVEKWQSSLDHVKRAADTAGRKAIRKVEDAVYRNVMTRIAPYYFDNALVSANIQ
                     RVRSQEEFVFEVNINDDDLKSAVQDHLSSFEALDRWHVDTERETDVAEAAEGVDPPKQ
                     REDAKSTTN"
gene                 5278. .6459
                     /gene="VNG2307C"
CDS                  5278. .6459
                     /gene="VNG2307C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng2307c"
                     /protein_id="AAG20418.1"
                     /db_xref="GI:10581719"
                     /translation="MVGAEAATFVVAAVASLFMAWAIGAGSSGSTPFAPAVGANAITV
                     MRAGFVVGILGLAGAVFQGANVSEAVGGDLIVGMPLSPLAATIALSTAALLVAAGVFA
                     GYPIATAFTVTGAVVGTGLAMGGGPAWPKYQQIATLWVLTPFVGGGTAYVTARSLRDT
                     AHERVLVPALAGGVGAITANIEFAFLGGPDAARSIAWSLASVAGTPSLVGRGAATLAV
                     AAAVAGALYCDSAGNASRGQRHFLLALGGLVAFSAGGSQVGLAVGPLLPLIGGEIPLP
                     YVLSGGGVGLLVGSWTGAPRMIKALSQDYSSLGPRRSIAALIPSFAIAQSAVFFGIPV
                     SFNEIIVSAIIGSGYAAAGGGGGGVSGRKMAVTVLAWVASLVVAFAASYGVFAAVSAV
                     M"
gene                 complement(6474. .7847)
                     /gene="hlp"
                     /note="VNG2308G"
CDS                  complement(6474. .7847)
                     /gene="hlp"
                     /note="Hlp"
                     /codon_start=1
                     /transl_table=11
                     /product="hemolysin protein"
                     /protein_id="AAG20419.1"
                     /db_xref="GI:10581720"
                     /translation="MGISPPLTAVSAFGTDVSGNAFIGAGVATLAVLMVLSAFFSSSE
                     IAMFSIPRHRIDALVEEGRSGADAVAALKSDPHRLLITILVGNNLVNIAMSSIATGLF
                     ATSMSQGKAVLAATFGVTALVLLFGESAPKSYAVENSESWALSIAKPLQWSERFLYPL
                     VVLFDYLTRAVNQFTGGRAAIESSYVTRSEIQDMIKTGEREGVIEEDEREMLQRIFRF
                     NNTIAKEVMTPRLDVVAVSKTDTIEAAIQTCTQAGHERVPVYDGELDNVIGVVSLEDL
                     VRESLYGETEDAELDDLIEPTLHVPESKNVDDLLQEMQDERVQLVVVIDEFGTTEGLL
                     TAEDITEEIVGEILDGGEDLPIDSVDEDTVRVRGEVNIEEVNEALNIDLPEGEEFETI
                     AGFIFNRAGRLVEEGEAFRFEDIELTVEHVENTRIMKARVQRLDSAAEADANAEVLTA
                     DSTGNGA"
gene                 7954. .8244
CDS                  7954. .8244
                     /gene="VNG2310H"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng2310h"
                     /protein_id="AAG20420.1"
                     /db_xref="GI:10581721"
                     /translation="MSAPAAPPRPPHSDAHITLYRLQACPFCERVVNRLEELGLAYQS
                     RYVEPMHSERDAVKRIVGARTVPAIVDDETGVAMAESANIVAYLDATYGGGE"
gene                 8244. .8759
                     /gene="VNG2311H"
CDS                  8244. .8759
                     /gene="VNG2311H"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng2311h"
                     /protein_id="AAG20421.1"
                     /db_xref="GI:10581722"
                     /translation="MELDFDVVDLPAADHVAAGDDAPDFIRPLVTDEYWADVALADLL
                     AEGPVVLVFTPMDGAFPATDMYSELRERGVHERAQVVGVSISTPYEHKTTIKDRGIQE
                     FTGLFSDPQAGVAEAYGVEHDPDGMAGVVEHRPATVVIGADGTVEYAWVAQQWPEFPA
                     YDELEAALDEL"
gene                 8846. .9442
                     /gene="VNG2312C"
```

```
Query Match              6.6%;  Score 130.4;  DB 1;  Length 9825;
Best Local Similarity   50.6%;  Pred. No. 4e-29;
Matches  426;  Conservative    0;  Mismatches  401;  Indels   15;  Gaps    4;

Qy    1126 atcccgcttggacaggaagagaaacagctgactggcgcgaatgacaagtttgtcagggag 1185
           | |||| | ||    ||||||     ||| ||  |||||||   | | |   ||   |||
Db    2552 ACCCCGATGGGTGGTGAAGAGGTCACGCTAACGCGCGCGAACCGCGACTACCTCGACGAG 2611

Qy    1186 catacgatatcaatacctgaagagattcaatatgctacagaagacggcgtgatggtgaac 1245
           |   | | || |  || ||   | |       | |       |  ||||  |   |  |
Db    2612 CGCGCCGTCTCCAAGCCCGAGTCGGTGTGGATCGATGGGCCCGCGGGCGACA---TCCAG 2668

Qy    1246 ggctggctgatgaggcctgcacaaatggaaggtgagacaacatatccacttattcttaac 1305
           || ||| || | | |||  |  | | || |||   || ||||| ||  |  | |||
Db    2669 GGGTGGGTGCTCACGCCCCCGGAGTTCGACCCCGAGGAGACGTATCCGCTGGTCGTGAAC 2728

Qy    1306 atacacggcggtccgcatatgatgtac------ggacatacatattttcatgagtttcag 1359
             | |||||||| |||||  ||||||          |    ||     |  || ||||| |||
Db    2729 GTCCACGGCGGCCCGCACCTGATGTGGTCGGCGGCCGGGACGATGTGGCACGAGTTCCAG 2788

Qy    1360 gtgctggcggcgaaaggatacgcggtcgtttatatcaatccgagaggaagccacggctac 1419
           |||  |||||     || |||| |||  | |      || ||  | ||        || |||
Db    2789 ACGCTTGCGGCCCGCGGGTACGTGGTGTTCTGGTGTAACCCCCGGGGCTCGACGGGGTAC 2848

Qy    1420 gggcaggaatttgtgaatgcggtcagaggagattatgggggaaaggattatgacgatgtg 1479
           || ||||| |    ||  ||| |||  | | || |  || |            |  || ||
Db    2849 GGCCAGGAGTGGCTGCGTGCCGTCGAACGCGACTGGGGCGCGGTCACGCTCGCGGACGTC 2908

Qy    1480 atgcaggctgtggatgaggctatcaaacgagatccgcatattgatcctaagcggctcggt 1539
           |||    | ||||  | ||  | | | || ||     |  | ||  |  |
Db    2909 ATGGCCGGCGTGGCGGCGGTCACCGACCGCGA---CTACGTGGACGCCGACAACGCGTTC 2965

Qy    1540 gtcacgggcggaagctacggaggttttatgaccaactggatcgtcgggcagacgaaccgc 1599
           ||||||||||| |||| ||| || || |||||| ||||   ||||| || ||  ||
Db    2966 ATCACGGGCGGGAGCTTCGGCGGGTTCATGACCGGCTGGGCGGTCGGCCACACCGACGCG 3025

Qy    1600 tttaaagctgccgttacccagcgctcgatatcaaattggatcagctttcacggcgtcagt 1659
           ||    |  || ||  | ||||||    | |   |    |   | ||  |||||   |
Db    3026 TTCGCCGGCGCGGTCGCGCAGCGCGGCGTCTACGACCTCACGGGGTTCTACGGCACGACG 3085

Qy    1660 gatatcggctatttctttacagactggcagcttgagcatgacatgtttgaggacacagaa 1719
           ||   |  |    || ||  ||  | || ||    |||    |   |||||  |
Db    3086 GACGCCTACGCGCTCGTT---GAGGGTGAGTTCGACACCGACCCCGTGCGGGACAACGCG 3142

Qy    1720 aagctctgggaccggtctcctttaaaatacgcagcaaacgtggagacaccgcttttgata 1779
             || ||||  | ||| ||     || | |    ||| || || ||     || |
Db    3143 TTCCTGTGGGCGCAGTCCCCGGCGGCGCACACCGACGCCGTCGACACCCCCACGCTGCTC 3202

Qy    1780 ctgcatggcgagcgggatgaccgatgcccgatcgagcaggcggagcagctgtttatcgct 1839
           ||||   |||   ||  ||||    |||    |    |||||||  | | |
Db    3203 GTGCACTCGGAGGCCGACTACCGGACGCCGGCGAACACCGCGGAGCTGTTCTACCGGCTG 3262

Qy    1840 ctgaaaaaaatgggcaaggaaaccaagcttgtccgttttccgaatgcatcgcacaattta 1899
           ||||| |||   |||   || ||| ||| || || | | |||    |    ||| || |
Db    3263 CTGAAGAAACACGGCGTTGACACCCGGCTGGTTCGGTATCCCGGCGAGGGCCACGATCTC 3322

Qy    1900 tcacgcaccggacacccaagacagcggatcaagcgcctgaattatatcagctcatggttt 1959
           || ||  |||| | ||  | ||   | || | ||| |  |    |||    | |||||
Db    3323 TCCCGGTCCGGTGATCCGGGCCACGTGGTCGACCGCATCGAGCGCATCGCTGCCTGGTTC 3382

Qy    1960 ga 1961
           ||
Db    3383 GA 3384
```

Search completed: September 26, 2001, 08:23:28
Job time: 19340 sec